CONSERVATOIRE NATIONAL DES ARTS ET METIERS

INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION

MEMOIRE

pour obtenir le

DESS en Sciences de l'information et de la documentation spécialisées

présenté et soutenu par

LAURÈNE MANSUY

le 16.11.2004

ETUDE DES LANGAGES DOCUMENTAIRES DES BANQUES D'IMAGES EN LIGNE : LE CAS DE L'AGENCE CIRIC

Jury composé de Claude GANTER , Cécile KATTNIG

Cycle supérieur Promotion XXXIV

Remerciements

Je tiens à remercier l'équipe du Ciric pour son accueil chaleureux et plus particulièrement Claude Ganter qui m'a soutenue et aidée pendant l'élaboration de ma mission, ainsi que Cécile Kattnig pour ses conseils avisés.

Résumé

Ce mémoire a pour objectif de décrire les méthodes de traitement documentaire de l'image photographique dans les agences de presse et d'illustration en ligne, à l'heure des nouvelles technologies. Il s'appuie pour cela sur l'étude du fonctionnement documentaire d'une agence de presse et d'illustration spécialisée, et sur la comparaison de plusieurs sites de banques d'images en ligne.

MOTS-CLÉS
banque d'images , document photographique, gestion électronique de documents, indexation, langage documentaire, SGBD, photothèque, traitement image

ENGLISH DESCRIPTORS
image databank, data base management system, image processing, indexing language, document retrieval system, photographic document, pictures library

Table des matières

Remerciements **2**

Résumé **3**

Table des matières **4**

Introduction **8**

LE TRAITEMENT DOCUMENTAIRE DE L'IMAGE FIXE **10**

I. Image fixe et photographie **11**

1. La photographie 11
2. La photographie et la presse 12
3. Emergence de la photographie numérique 12

II. La lecture de l'image **13**

1. Approche sémiologique de l'image 13
2. La polysémie de l'image 14
3. La spécificité de la recherche iconographique 14

III. Le traitement documentaire **16**

1. La numérotation, le légendage, l'estampillage, la numérisation et le stockage 16
2. le catalogage, l'analyse de l'image, le choix des langages documentaires 20

ETUDE DE L'EXISTANT 28

I. L'agence de presse et d'illustrations CIRIC 29

1. Historique de l'agence 29
2. Typologie des usagers du CIRIC 29
3. La base de données CIRIC 30
4. Le lexique de la base 31

II. le transfert de la base vers le nouveau logiciel choisi 35

1. Nettoyage du lexique 35
2. La prise en main du nouveau logiciel 36
3. Applications de nouvelles règles d'écriture 38
4. Améliorations du langage documentaire 40

III. Application de la méthode du Benchmarking pour la construction de l'interface web 43

ETUDE COMPARATIVE DU TRAITEMENT DOCUMENTAIRE DANS LES AGENCES PHOTOGRAPHIQUES EN LIGNE 47

I. Les fonds d'images en France 48

II. L'évolution du marché de la photographie de presse et d'illustrations 50

III. Quelques logiciels spécialisés dans l'image 53

1. Principales fonctionnalités des logiciels 53
2. Grille d'évaluation de 5 logiciels 54

IV. Analyse des fonctionnalités documentaires de plusieurs banques d'images en ligne 57

1.Grilles d'évaluation des banques d'images 57

2. Résultats et surprises 61

V. Tendances 62

1. L'émergence de portails 62

2. L'indexation automatique 63

Conclusion 65

Bibliographie 67

Annexes 74

Annexe 1 : Plan de classement proposé par l'agence Lumière du monde 75

Annexe 2 : Interface de recherche sur l'ancienne base de Ciric (DFI) 76

Annexe 3 : Interface de recherche sur l'ancienne interface web de Ciric (SACD) 77

Annexe 4 : Interface web actuelle de CIRIC (page d'accueil) 78

Annexe 5 : Interface web actuelle de CIRIC (niveau 1) 79

Annexe 6 : Interface de recherche par mots-clés de l'agence Roger Viollet 80

Annexe 7 : Interface de recherche de l'agence Magnum 81

Annexe 8 : Affichage d'une photo avec notice documentaire et mots- clés associés, agence CIRIC 82

Annexe 9 : Affichage des résultats d'une recherche avec mise en ligne du thésaurus, agence Urba Images 83

Annexe 10 : Portail d'ImageForum..... 84

Annexe 11 : Exemple d'indexation d'images par le contenu (Image Seeker) 85

Introduction

Nous assistons depuis quelques années à une inflation de la place de l'image dans notre société. L'image, fixe ou animée, est devenue un support nécessaire à toute communication. Le mode d'accès aux images a été considérablement transformé par la technologie numérique et l'utilisation du réseau internet. Beaucoup de banques d'images ont numérisé leurs fonds ces dernières années et se sont dotées d'un accès sur le web.

De nombreuses photothèques vendent aujourd'hui leurs documents en ligne, qu'elles concernent l'actualité ou l'illustration. Le monde de la presse, de l'édition et du multimédia a vu ses pratiques professionnelles totalement remaniées. Ces photothèques ont dû envisager des solutions pour optimiser leur système documentaire afin que leurs clients des secteurs de la presse, de l'édition ou de la publicité soient quasi-autonomes dans leurs requêtes et la satisfaction de celles-ci, ceux-ci n'ayant dans la plupart des cas aucune compétence en informatique documentaire. Le problème de l' organisation de ces images s'est donc posé le via la conception de systèmes permettant une sélection rapide et pertinente des images recherchées ainsi qu'une diffusion efficace.

J'exposerai dans une première partie les spécificités du traitement documentaire de l'image fixe et je décrirai la transformation que les nouvelles technologies ont apporté pour les agences photo en terme de stockage, de consultation et de diffusion de ces images.

Puis j'étudierai cette problématique à l'échelle de l'agence photographique Ciric, spécialisée dans les religions et les questions sociales, qui s'est dotée en 2004 d'un nouveau logiciel de gestion documentaire d'images et qui a dû adapter sa base de données à ce nouvel outil.

Enfin dans une troisième partie je dresserai un bref état des lieux des fonds d'images en France aujourd'hui. Je m'attacherai à décrire la modification du marché de l'image de presse et d'illustration dans le contexte actuel de rachat et de concentration, ainsi que les répercussions que cela suscite pour le traitement documentaire. Puis je comparerai quelques sites de banques d'images en ligne afin d'évaluer les diverses méthodes de présentation, d'analyse et d'acquisition de l'image sur internet.

Je tenterai de mettre en évidence les différences et les points communs des pratiques documentaires de ces quelques sources et m'interrogerai sur l'avenir de ces usages.

LE TRAITEMENT DOCUMENTAIRE DE L'IMAGE FIXE

I. Image fixe et photographie

Une image fixe est un document iconographique immobile. Il peut s'agir d'une peinture, d'un dessin, d'une photographie, d'une bande dessinée, d'une caricature ou d'une carte géographique.
On distingue l'image fixe de l'image animée. Celle–ci est un document audiovisuel, qui est composé d'images (plusieurs images fixes associées qui constituent une image en mouvement) mais qui peut aussi inclure des sons.

Je centrerai mon travail sur les images photographiques dans les agences de presse et d'illustration.

1. La photographie

Une photographie est un procédé technique permettant d'obtenir une image sur une surface sensible. Son invention a été officiellement annoncée le 19 août 1839 à Paris, sous la forme primitive du daguerréotype. Mais quatre hommes ont participé à l'invention de la photographie tout au long du 19ème siècle :
Nicéphore Niepce avait commencé dès 1816 ses recherches et expérimenté divers supports : papier, verre, pierre, métal, enduits de matières photosensibles (en particulier le bitume de Judée) exposés à la lumière de la chambre noire.
Louis-Jacques Mandé Daguerre mit au point une méthode ingénieuse mais complexe qu'il exploita sous son nom, le daguerréotype : une image positive, unique et non reproductible obtenue sur plaque de cuivre argentée rendue sensible à la lumière par une vapeur d'iode. Cette invention en déclencha deux autres : l'anglais William-Henri Fox-Talbot travailla sur le support papier en utilisant un papier imprégné de sels d'argent qui, soumis à la lumière, garde l'empreinte des objets posés à sa surface (les Photogenics Drawings). Il obtint le premier négatif en 1841 : le calotype, à partir duquel plusieurs exemplaires peuvent être tirés. Et Hippolyte Bayard mit au point vers 1839 un procédé à mi-chemin entre le calotype et le daguerréotype qui permit d'obtenir une image unique sur papier mais sans passer par un négatif.
De nombreux autres créateurs poursuivirent l'amélioration des procédés en rationalisant la méthode de fabrication des négatifs papier, réduisant ainsi le temps de pose, en remplaçant le négatif papier par du verre ou en améliorant la qualité des chimies, avec l'émulsion aux sels d'argent par exemple. La chambre noire avait fait son temps : Georges Eastman créa le premier appareil photographique Kodak en 1889. La photographie couleur apparut en 1903 avec l'invention des frères Lumière. L'autochrome se généralisa. Dans les années 30 beaucoup de découvertes portant sur la qualité des négatifs film furent réalisées, donnant naissance aux pellicules Kodachrome, Ektachrome, au Polaroïd.

2. La photographie et la presse

La photogravure permet à partir de 1878 l'illustration d'ouvrages, de journaux et de magazines avec un tirage en nombre, qui va se développer grâce aux machines rotatives et à l'offset. C'est le 4 mars 1880 qu'on trouve la première « Une » photographique dans le New York Daily Graphic. En France le premier quotidien illustré est crée en 1909 : l'Excelsior. En 1914 a lieu la première transmission d'un cliché à distance par radio puis par ligne téléphonique grâce à l'invention du bélinographe.
Ce procédé a été utilisé jusque dans les années 90 et a été remplacé depuis par la transmission numérique. Les satellites, les fibres optiques et le laser ont largement accéléré les vitesses de transmission et augmenté les possibilités de volume transmis.
La photographie s'est vraiment imposée dans la presse dans les années 20 / 30 grâce à l'évolution des techniques de reproduction et la mise au point d'appareils photographiques (Leica en 1925, Rolleifleix en 1929). Des agences photographiques se sont donc créées, regroupant des photographes indépendants pour alimenter la presse alors en pleine expansion et organiser la diffusion des images : Delphot à Berlin en 1928, Alliance Photo, Rapho et Magnum à Paris dans les années 30 / 40.
L'évolution s'est faite du tirage argentique au format numérique ces dernières années.

3. Emergence de la photographie numérique

L'image numérique est le résultat soit d'une prise de vue numérique, soit d'une scannérisation d'un document iconographique. Elle a bouleversé les chaînes de production, photographique et éditoriale.
De 1970 à 1990 environ, le développement des mémoires optiques a permis la généralisation de l'archivage de fonds d'images. Le vidéodisque Laservision, le disque optique numérique (DON) ont initié le gravage numérique pour de multiples collections d'images appartenant aussi bien au monde de l'entreprise, des musées ou des photothèques indépendantes (musée départemental Albert Kahn, EDF, Institut du Monde Arabe, Agence Gamma …) Depuis une quinzaine d'années la standardisation des mémoires optiques, la normalisation de la compression et de la transmission d'images ainsi que la puissance grandissante des ordinateurs ont modifié les pratiques de travail ; les grandes agences photographiques ont incorporé les images dans leur chaîne éditoriale numérique, intégrant et diffusant leur fonds sur leurs réseaux pour leurs abonnés. Les musées ont mis en valeur leur collection par des éditions multimédia et des sélections d'images présentées sur internet.

II. La lecture de l'image

1. Approche sémiologique de l'image

La sémiologie c'est la «science qui étudie les systèmes de signe», selon Le Petit Robert. Cette discipline tire ses origines du XVIIe siècle. On peut identifier trois «pères fondateurs» de la sémiologie (encore appelée «sémiotique»): le britannique Locke (1632-1704), l'américain Peirce (1839-1914) et le suisse Saussure (1857-1913).

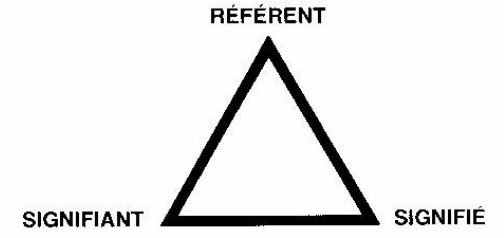
La sémiologie « s'intéresse à la nature de la relation signifié/signifiant, au fonctionnement du signe dans son ensemble (structural et contextuel) qu'il soit linguistique ou non » (L'image et les signes, p.15) [10]

Le système de signes le plus connu et le plus utilisé consciemment par les humains, est la langue, les mots. Mais un grand nombre d'autres codes sont utilisés couramment. Pensons aux signaux de circulation, au langage gestuel des sourds-muets, au Braille des aveugles … Les images de façon générale constituent aussi un système de signes.

Un des premiers à avoir pratiqué l'analyse sémiologique des messages visuels est Roland Barthes dans les années 60. Barthes considérait les images comme des combinaisons de signes et a montré que tout système de signes se mêle de langage verbal : il a constaté en effet que dans toute image (cinéma, télévision, publicité, photo de presse…) le langage verbal double la substance visuelle et entretient alors, dans presque tous les cas, un rapport structural avec le message visuel. Pour Barthes, " percevoir ce qu 'une substance signifie, c 'est fatalement recourir au découpage de la langue : il n 'y a de sens que nommé et le monde des signifiés n 'est autre que celui du langage..."(Rhétorique de l'image, p.45) [11]

Le sémiologue allemand Max Bense (1971) résume sous une forme schématique la nature tridimensionnelle d'un signe selon qu'il est en relation avec :

- l'objet qu'il représente (ou référent)
- le signifié qu'il produit (la représentation mentale qu'un individu se fait du référent au contact du signifiant)
- le signifiant qu'il est (l'aspect sensible du signe)



Selon le type de rapport qu'un signe entretient avec son référent, le signe sera un **icône** (ressemblant à l'objet), un **indice** (ayant une relation de «contiguïté naturelle») ou un **symbole** (fondé sur une convention arbitraire).

L'information portée par un signe (sa signification) correspond au lien dialectique permanent unissant le signifiant et le signifié. Mais cette signification même est ambivalente : s'il existe une signification dénotative renvoyant au niveau référentiel d'une image, il existe aussi une signification connotative renvoyant à son interprétation.

2. La polysémie de l'image

L'image fournit un grand nombre d'informations, peut avoir de nombreuses significations tant dénotatives que connotatives, et donc se prêter à de multiples interprétations.

Depuis quelques années la sémiotique visuelle a évolué et s'attache maintenant à reconnaître l'importance du contexte dans laquelle l'image est décryptée : « une image est en effet perçue et lue en fonction de la manière dont elle est produite, diffusée et reçue » (l'invisible dans l'image, p.27) [8]

Les niveaux de lecture d'une image dépendent intégralement de l'identité culturelle et sociale de la personne qui reçoit cette image et du contexte d'usage de cette image. L'interprétation dépend non seulement de la sphère culturelle de la personne qui analyse l'image mais aussi des relations de sens que cette personne établit entre l'image et sa sphère culturelle. C 'est ce que Henri Hudrisier appelle « le pouvoir évocateur de l'image » (L'Iconothèque, p.164) [30]

La polysémie de l'image peut poser de nombreuses problèmes dès lors qu'il s'agit d'analyser, au sens documentaire du terme, une image. Comment faire apparaître tous les concepts présents dans une image ? Est-il nécessaire que tous ces concepts soient décrits ? Nous verrons plus loin à quel point cette analyse est liée à la notion de « contexte » ou d' « environnement » et dépend de l'usage prévu d'une image.

3. La spécificité de la recherche iconographique

L'iconographe a pour fonction de fournir des images à la presse, l'édition, la publicité ou les institutions culturelles. Il lui faut donc répondre conjointement à une demande rédactionnelle (quelle image pour quel propos ?) et à une demande esthétique (quelle image pour quel support ?).

La recherche iconographique ne résulte absolument pas des mêmes méthodes qu'une recherche textuelle. Contrairement à une recherche textuelle, beaucoup plus ciblée, une recherche iconographique ne vise quasiment jamais à trouver « l'image » qui correspond exactement à ce qu'on cherche, mais un ensemble d'images se rapprochant d'une requête.

Parler d'image « juste » ne revient surtout pas à dire qu'il n'existe qu'une image possible en réponse à une demande, mais qu'il existe celles qui peuvent être retenues parmi tant d'autres. Cela nécessite de cultiver à la fois une approche objective de l'information (donc une vraie réflexion sur celle-ci) et une approche subjective ouverte à l'interprétation personnelle de chacun.

La recherche d'images dans son essence même privilégie le phénomène du bruit au silence : il sera plus facile de choisir une image en la comparant à d'autres images, ce que proposent aujourd'hui de nombreux logiciels documentaires avec la visualisation par mosaïques.

Avec la multiplication des sites consacrés aux images, l'iconographe a accès à un très grand nombre de photos et le seul moyen de ne pas s'égarer est souvent lié à la qualité d'indexation du fonds.

III. Le traitement documentaire

Le traitement documentaire de l'image fait référence aujourd'hui, avec l'avènement des nouvelles technologies, à toutes les étapes de la chaîne documentaire subies par l'image avant d'être visible sur un écran d'ordinateur. Il concerne à la fois la gestion documentaire et l'exploitation de ces objets.

La gestion des collections commence par une politique d'acquisition qui détermine le genre, le type et le nombre des documents qui seront proposés aux usagers / clients. Cette opération qui se classe au début de la chaîne documentaire permet la création et l'accroissement du fonds par la collecte, l'acquisition, et la sélection. Dans le cadre des photothèques cette étape concerne le choix éditorial de chaque agence et prend en compte la problématique du droit d'auteur ; nous n'en ferons pas mention ici.

Les premières étapes à prendre en compte sont :

1. La numérotation, le légendage, l'estampillage, la numérisation et le stockage

Plusieurs tâches liées au traitement matériel doivent être effectuées pour l'intégration des images dans une base de données, quelles soient réalisées manuellement ou par le système de gestion informatisée.

En premier lieu il faut numéroter et légender les images.

La numérotation

Chaque reportage reçoit un numéro d'inventaire, ainsi qu'un numéro propre à chaque photo. Ce numéro est essentiel car il identifie une photo et lui confère son caractère d'unicité. Le principe de numérotation est généralement lié à une logique de rangement original à chaque fonds : il peut porter sur un code support / format ou sur un code thématique renvoyant à un plan de classement dans une photothèque. Il peut aussi s'appuyer sur un ordre chronologique par photographe. Il peut être aussi exclusivement chronologique.

Le légendage

Il est obligatoire en regard de la loi sur le droit d'auteur et il est de toute façon indispensable pour retrouvrer une photo. Il est élaboré avec le photographe à son retour de reportage, et est structuré afin d'en faciliter la lecture et la recherche plein texte sur la base de données. La légende ne doit pas mentionner ce qui est évident et immédiatement repérable sur l'image mais ce que l'image ne dit pas et qui compte pour sa bonne lecture.

L'estampillage

Les images lorsqu'elle sont sous forme de papier ou d'ektas sont numérotées et légendées sur le support, qu'il s'agisse de l'original ou de reproductions.

Les mentions à y inscrire sont :

- Le nom du photographe
- Le copyright et l'adresse de l'agence
- La légende et la date
- Si possible le numéro du cliché dans la photothèque

Pour les prises de vue numérique ou les images numérisées, les données sont associées en fichier texte comme nous le verrons avec le format IPTC.

Aujourd'hui pour les images accessibles sur site internet ou sur CD Rom on estampille la photo avec un marquage filigrane ou watermark. Ces marquages numériques sont le plus souvent constitués par un logo ou une mention obligatoire pour la protection des droits d'auteurs. On peut également tatouer une image afin de la rendre inutilisable commercialement, évitant ainsi le piratage.

Il faudra également numériser les photos, éventuellement les restaurer et les retoucher.

La numérisation

L'image numérique est une image sous forme de fichier informatique. Pour numériser l'information, on enregistre non pas le signal continu, comme c'est le cas lors d'un enregistrement analogique, mais la valeur de ce signal en un nombre limité de points de l'image, appelés en anglais les « picture elements » ou « pixels ». L'image est donc divisée en un quadrillage à l'intérieur duquel sont prélevés les pixels. Plus le quadrillage est serré, plus la fidélité à l'image d'origine est grande. Le nombre de pixels par unité de longueur constitue la résolution de l'image et s 'exprime en dpi (dots per inch ou points par pouce, 1 pouce étant l'équivalent de 2,54 cm). La résolution est définie par l'appareil de numérisation choisi (scanner à plat, scanner à diapositives, scanner rotatif, appareil photo numérique).
La résolution s'applique lors de la numérisation mais aussi lors de la consultation sur écran informatique et lors de l'impression. La résolution moyenne est de 72 dpi.

Une fois produit, le fichier numérique occupe un certain volume sur le support de stockage, que celui-ci soit un disque dur, un CD Rom ou autre. Ce volume s'exprime en octets.

On a vu que l'image est décomposée en fragments élémentaires appelés pixels. Face à l'importance des volumes (il est rare de disposer d'un support capable d'archiver plusieurs centaines de millions d'octets) , il faut réduire la taille de ces fichiers : la compression proprement dite permet, avant transmission, de réduire le nombre d'éléments binaires nécessaires à la description du document grâce à l'enregistrement de l'image sous un format spécial : JPEG, TIFF, BMP ou GIF.

Ces formats permettent d'utiliser environ vingt fois moins de place, ce qui permet de mettre environ 15 000 images sur un disque dur d'un giga octet. Mais il faut savoir que plus les informations sont compressées, moins la qualité et la luminosité sont bonnes.

Il existe deux grands formats d'images numériques : les bitmaps et les vectoriels. Les formats bitmaps sont davantage destinés aux photographies alors que les formats vectoriels conviennent mieux à des modélisations en 3 D. Ceux-ci sont constitués de multiples objets géométriques (lignes, ellipses, polygones, entre autres formes) construits grâce à une séquence de commandes ou de formulations mathématiques pour tracer des lignes et des formes. Les graphiques vectoriels sont créés et stockés en utilisant un format ouvert tel que Scalable Vector Graphics (SVG), un langage XML permettant de décrire de tels graphiques.

Dans le format bitmap destiné aux photographies on distingue :

- Le BMP
- Le TIFF (Tagged image file format)
- Le GIF (Graphics interchange format)
- Le JPEG (Joint photographic expert group)
- Le PNG (Portable network Graphics)
- L'EPS (Encapsuled Poscript)

Les plus fréquemment utilisés sont TIFF, GIFF, JPEG (Images numériques : quels format, quelles caractéristiques, pour quels usages ?, p.383) [16]

Le TIFF permet d'enregistrer des images avec ou sans compression.
Le GIF est l'un des plus courants mais il code uniquement 256 couleurs par pixel.
Le JPEG est la représentation d'une norme ISO, conçu par le groupe expert en technologie de l'image fixe. C'est à la fois un format de fichier et un standard de compression.

« Les avantages de la numérisation sont nombreux : grande souplesse du traitement, permanence de l'information, support inusable, baisse des coûts, qualité de la restitution de l'image, possibilité de transmission à distance, grande variété de solutions de consultation - du micro-ordinateur au gros serveur - et excellente impression sur papier, proche du tirage argentique, à partir d'une imprimante laser. Tous ces éléments ont permis de générer la

création d'importantes banques d'images." (Les images dans les bibliothèques, p.139) [25]

Si la compression a permis de résoudre le problème du stockage et de la diffusion des fonds iconographiques, il faut néanmoins se rappeler que l'image est toujours fortement consommatrice de mémoire sur les serveurs, d'où le problème de flux qui ne manque pas de se poser lors de téléchargements.

La retouche s'effectue parfois sur les images scannées pour pallier les problèmes de colométrie ou de flou, à l'aide de logiciels de traitement d'images , dont le plus répandu est Adobe Photoshop.

Le stockage et la visualisation

L'image ainsi numérisée est stockée sur un support informatique (ordinateur, CD Rom , DVD, …) dans le but d'être visualisée en local, à distance ou reproduite.

La visualisation des images s'effectue selon plusieurs modes possibles :

- image vignette (environ 230 X 360 pixels) correspond à la basse définition ; elle est conçue pour la transmission a distance.
- image plein écran (530 X 640 pixels) correspondant à la moyenne définition ; elle permet l'accès en ligne.
- image haute résolution (2000 X 3000 pixels) correspondant à la plus haute définition ; elle est indispensable dans le secteur de la presse.

Un CD Rom dont la capacité est de 650 MO (méga octets) peut stocker environ 400 images en haute définition comprimées.
Les images peuvent aussi être stockées sur un DVD dont la capacité est de 7 à 20 fois supérieure au CD Rom.

L' avènement du numérique a réduit le stockage et le classement puisque dans la plupart des cas aujourd'hui l'original, s'il n'est pas déjà numérique, va être scanné et intégré dans la base de données, évitant dans certains cas la consultation de l'original. Il a facilité la diffusion électronique via internet, extranet ou intranet avec une plus grande vitesse d'accès aux documents et des affichages variés. Enfin il a rendu possible une manipulation simplifiée et plus légère.

Il faut malgré tout être vigilant sur les problèmes de pérennité des supports informatiques dans leur ensemble. Deux types de difficultés peuvent entraver l'accès continu aux supports numériques :

- la détérioration physique ou l'endommagement du support lui-même,
- le changement technologique conduisant à l'obsolescence de l'infrastructure logicielle et matérielle nécessaire pour accéder au support.

On distingue enfin tout ce qui a trait au traitement intellectuel de l'image, à savoir :

2. le catalogage, l'analyse de l'image, le choix des langages documentaires

Le catalogage

Le descriptif bibliographique de l'image s'effectue selon une norme de catalogage propre à l'image fixe et l'indexation est toujours réalisée à partir d'un langage documentaire spécifique au fonds concerné.

Règles de catalogage

La présentation de l'image comporte toujours des indicateurs de sa légende, indispensable à sa lisibilité : indicateurs de source, de lieu, d'auteur, date, titre, n° de référence …

La description bibliographique de l'image fixe est définie dans la norme AFNOR NZ 44-077 : « documentation – catalogage de l'image fixe – rédaction de la description bibliographique » élaborée dès 1985 par une commission d'experts animée par Denis Bruckmann, conservateur au Département des Estampes et de la Photographie, et publiée en 1997. [20] Elle a pour origine les recommandations internationales de la description des documents autres que livres (NBM : Non Book Materials).

Elle permet de choisir entre un catalogage par lots ou un catalogage image par image.

Elle comprend 8 zones :

- Zone 1 : titre (entre crochets) et mentions de responsabilité (auteurs)
- Zone 2 : nature du support , comme état du dessin ou du tirage photographique
- La zone 3 n'est pas utilisée pour les photos
- Zone 4 : lieu d'édition, éditeur et date d'édition ; généralement c'est la date de prise de vue pour une photographie
- Zone 5 : description formelle concernant le support, le format, la notion de couleur ou noir et blanc, la technique employée, la couleur et le format
- Zone 6 : collections dont l'image est issue, dans un album ou un portfolio par exemple
- Zone 7 : zone des notes
- Zone 8 : zone des numérotations antérieures.

Chaque élément de description est introduit par une ponctuation précise .
Chaque zone est séparée par un .- (point tiret espace).

Chaque fonds d'images élabore un format de description qui lui est propre prenant en compte ses besoins et sa spécificité. Il adapte la notice descriptive en fonction de la spécificité et de la structure de son établissement, de la nature de ses documents, de ses types d'usagers. Bibliothèques, titres de presse, musées,… ne choisiront pas la même mise en forme de cette notice.

Une bibliothèque choisira le format MARC (Machine Readable Cataloguing), le format d'échange utilisé par les bibliothèques ; le format MARC est issu d'US-MARC créé par la Bibliothèque du Congrès aux Etats-Unis dans les années 60. Il existe plusieurs formats MARC selon les pays mais le format international UNIMARC a été créé pour normaliser les échanges internationaux.

D'autres fonds d'images utiliseront le Dublin Core, crée à Dublin (Ohio) en 1995 par plusieurs équipes impliquées dans la sémantique sur le Web . Ce format définit une liste précise de quelques métadonnées liées aux pages Web. Les 15 champs de métadonnées du Dublin Core sont appelés « éléments » ; la signification de certains d'entre peut être précisée à l'aide de « raffinements ».

Le secteur de la presse quant à lui intègre le format de transmission IPTC.

Le format IPTC (International Press Telecommunication Committee)

L' IPTC est une organisation internationale créée en 1965 pour développer et promouvoir des standards d'échange de données à destination de la presse et des médias nord-américains. Il s'agissait d'acheminer de façon simple des informations destinées à classer automatiquement des fichiers par les ordinateurs à l'aide de mots-clé présents dans les en-têtes de ces derniers.
Avec ce format les professionnels ont proposé un encapsulage des informations dans le fichier image lui-même. [13]

En association avec la NAA (Newspaper Association of America), l'IPTC a défini en 1992 un modèle global de données appelé IPTC-NAA Information Interchange Model (IIM). La structure est adaptée aux besoins d'échange de tous types de données dans la presse et les médias. Le modèle comprend un format pour les données et un protocole de transmission. Il est structuré en une centaine de champs balisés dont certains sont obligatoires. L'image est ainsi automatiquement transmise avec les informations la concernant.
Par le biais des classifications et de l'évolution des différents standards de description de page au format texte : HTML, SGML, Unicode, etc, c'est l'ensemble des données transitant par les ordinateurs qui peuvent être aujourd'hui identifiées grâce à ces meta-informations.

Reconnue par la plupart des grands éditeurs de logiciels et par l'ensemble des professionnels de la presse et du multimédia, cette norme internationale constitue actuellement, le véritable standard. Tous les logiciels de bases de données ne sont pas à fortiori compatibles avec les champs IPTC mais la plupart le sont : Phrasea, Ajaris-Pro, Port-Folio, Orphéa Studio, Armadillo, Adobe PhotoShop (depuis la version 5)…

Ce modèle est structuré en une centaine de champs qui sont regroupés en 6 blocs, en principe renseignés par le photographe ou l'auteur de l'image :

- Légende (de l'image, du texte)
- Mots-clés
- Catégories (classification à 2 niveaux hiérarchiques permettant à l'auteur, dans le contexte générique de la presse, de situer son image dans une thématique avant de la transmettre à distance)
- Crédits (auteur image, auteur texte, copyright)
- Origine (date de création, pays, ville, date et heure d'envoi)
- Compression (qualité, format du fichier, validation)

L'IPTC propose des listes pour les mots- clés et les catégories. La présence des données texte associées à l'image offre une plus grande sécurité , et la notice descriptive avec pré-indexation à partir des listes « catégories » et « sous-catégories » facilite le travail documentaire.

Ces informations à la fois techniques et documentaires sont de plus en plus indispensables, et l'utilisation des champs IPTC constitue un premier pas vers cette organisation et cette structuration de l'information qui doit être contenue dans l'image. Avec le développement des bases de données et celui de la prise de vue numérique, nous sommes de plus en plus confrontés à des supports virtuels qui devront être identifiés au risque de se retrouver avec des milliers de données informelles.

L'analyse de l'image

L'analyse, étape préalable à l'indexation, permet de traiter l'image dans ce qu'elle décrit et dans ce qu'elle porte comme information. Elle a pour but de présenter le document sous une forme capable de permettre une consultation ou un repérage plus facile. Sa lecture n'est pas universelle mais multiple comme nous l'avons déjà évoqué en première partie.

Pour la photographie on dénombre 3 niveaux d'analyse :

1. **L'analyse morphologique** renseigne sur le support (négatif, ekta..) le format (carré, rectangulaire), l'optique (grand angle,..), le cadrage (plan moyen, contre plongée…), la lumière.

2. **L'analyse descriptive** concerne le contenu informatif de l'image et permet de répondre par des termes concrets aux questions :

- Où : la situation géographique,
- Quand : la date du sujet représenté
- Qui : les personnes ou les animaux présents à l'image,
- Quoi : les objets représentés, l'action
- Comment : le contexte

La méthode QQOQC est appliquée partout car elle est très fiable.

3. **L'analyse interprétative** ou **connotative** s'applique à la restitution de la charge affective d'une image. Elle est délicate car elle ne repose que sur peu d'éléments concrets et dépend entièrement de notre perception personnelle face à l'image.

Certains fonds n'utilisent pas ou très peu ce niveau d'analyse, d'autres comme la publicité l'utilisent à outrance.

Tous les fonds iconographiques développent des systèmes d'analyse documentaire personnels en fonction de la spécificité de leur fonds et de leurs usagers.

Principes d'indexation de l'image

L'indexation « est un processus destiné à représenter, au moyen de termes ou indices d'un langage documentaire ou au moyen des éléments d'un langage libre, les notions caractéristiques du contenu d'un document (ressource, collection) ou d'une question, en vue d'en faciliter la recherche, après les avoir identifiés par l'analyse. Les combinaisons possibles des notions identifiées sont représentées explicitement (indexation pré-coordonnée) ou non (indexation post-coordonnée) en fonction du langage utilisé » (Vocabulaire de la documentation) [1]

L'indexation est définie depuis 1978 dans la norme AFNOR Z47-102 : « Principes généraux pour l'indexation des documents ».

L'indexation de l'image consiste à extraire les concepts les plus représentatifs de l'image et à traduire ces concepts selon un langage documentaire approprié à la source d'images. « L'indexation est une opération indispensable. Elle permet non seulement le tri et l'accès, mais aussi l'identification du sujet et l'exploitation de la polysémie de l' image. Elle est à distinguer du catalogage qui consiste à donner une référence bibliographique à l'objet. » (Comment indexer l'image fixe ?, p. 33) [24]

L'indexation de l'image peut bien sûr être éclairée par les données textuelles qui identifient le document mais elle ne saurait être ainsi entièrement explicitée. Lorsque les images sont des représentations à l'identique d'objets, l'indexation revient à une indexation des objets correspondants ; mais lorsqu'il s'agit de décrire une action, un mouvement ou une ambiance, il

revient alors au documentaliste de choisir des termes au plus proche de ce qu'il voit à l'image.

Exemple :
Une photo de l'agence Magnum datant de 2002 et représentant une femme venant d'apprendre la mort de son mari au Cachemire, dans le cadre des conflits entre l'Inde et le Pakistan, est indexée de la façon suivante :

Asiatique de l'Asie du Sud . Bouche ouverte . Bracelet . Conflit armé . Costume et tailleur . Détresse . Deuil . Femme 25 à 45 ans . Flou . Homme 25 à 45 ans . Hôpital . Jammu et Cachemire . Peur (atmosphère, ambiance, sensation) . Peur (expression) . Pleur . Rideau . Tissu imprimé . Toucher . Tristesse . Vêtement traditionnel . Violence

Toute approche d'indexation s'appuie sur une connaissance précise des centres d'intérêt des utilisateurs et du contexte d'utilisation de l'image et présuppose une définition préalable stricte de la politique d'indexation au sein d'un fonds d'images.

La diversité de ces approches explique la grande variété des méthodes d'indexation, qui n'ont à ce jour fait l'objet d'aucune norme spécifique. En revanche, ces méthodes sont généralement appliquées avec des logiciels standards et selon des règles grammaticales normalisées, et s'appuient sur des outils d'aide qui peuvent être de plusieurs sortes :

- manuel de traitement documentaire comprenant des listes d'autorité
- mode d'emploi pour l'utilisation du vocabulaire

Selon les lieux on distingue une indexation image par image, une indexation par reportage et une indexation par lots d'images homogènes. Le traitement image par image se généralise avec le développement des banques d'images.

Le phénomène de surindexation est fréquent et il est recommandé pour les banques d'images. Il consiste à élargir l'indexation d'un objet à sa classe générale. Ce procédé est parfois traduit par l'expression « couvrir par ».

Exemple :
Chez AKG une représentation d'une peinture du Titien représentant Charles V avec son chien sera indexée avec les termes : ANIMAL, ANIMAL DOMESTIQUE, CHIEN, DOGUE.

La question est bien sûr toujours de savoir jusqu'à quel niveau d'indexation aller, car trop de surindexation provoque du bruit à l'interrogation.

Comme pour l'indexation de textes, l'indexation d'images porte sur 4 formes de langages différents :

- Le langage libre

Le documentaliste crée souvent un descripteur au plus proche de ce qu'il voit à l'image. Les notions sont représentées par des mots ou des expressions sans aucune relation sémantique.

- Le langage contrôlé ou liste d'autorité

Les fonds d'images constituent des listes de termes appliqués à l'image fixe qui sont des listes spécifiques ou listes d'autorité. Elles sont relatives à la morphologie de l'image, à la description de classes d'âges de personnes, elles listent des lieux géographiques, énoncent des sentiments, ordonnent des actions, ... Cette indexation nécessite qu'il y aie eu un travail d'élaboration en amont. Les listes sont définies et enrichies avec des règles bien précises et les mots sont classés par ordre alphabétique.

- Le langage classificatoire

La classification c'est la mise en ordre intellectuelle des documents : on rassemble des notions en catégories, selon un principe hiérarchique qui va du général au particulier. Cette organisation par arborescence est symbolisée par des indices. Elles sont le plus souvent encyclopédiques comme la classification DEWEY ou la CDU, traditionnellement utilisées en bibliothèques.

On trouve aussi des classifications propres à l'image, comme la classification ICONCLASS, qui fut mise au point en langue anglaise aux Pays-Bas. (Les images dans les bibliothèques) [25]

Inspirée de la classification décimale DEWEY, celle-ci couvre la totalité des sujets, thèmes et motifs des représentations dans l'art occidental. Elle indexe toute image par une cote alphanumérique. Les grandes catégories sont les suivantes :

1. Religion et magie.
2. Nature.
3. L'être humain.
4. Société matérielle et culture.
5. Concepts et abstractions.
6. Histoire.
7. Bible.
8. Littérature.
9. Mythologie et histoire ancienne

Ce système semble très satisfaisant pour décrire de l'iconographie traditionnelle en histoire de l'art (peintures, sculptures, etc.) mais inadapté pour tous les autres objets.

Les classifications mises au point dans les agences photographiques favorisent l'accès direct aux images dans leur rangement physique, indiqué par un indice de classification. Leur inconvénient majeur tient au fait qu'elle réduisent l'indexation à une seule notion. Quelques classifications sont mises en ligne sur les banques d'images, souvent en tant que première aide à la recherche [annexe n°1]. On remarque d'ailleurs un retour vers la mise en place de classifications sur les extranet et les intranet des grandes entreprises, dans le but de permettre un meilleur accès aux documents, ceux-ci étant de plus en plus nombreux, et leur repérage par conséquent de plus en plus difficile. La recherche peut ainsi s'opérer selon deux modes :
- le mode classificatoire
- le mode « mots clé » http://savoirscdi.cndp.fr/rencontrelyon/kattnig/ - haut

- Le langage combinatoire ou thesaurus

Le thésaurus permet de normaliser les termes utilisés servant à l'indexation et à la recherche. Il est construit selon les normes AFNOR Z 47-100, 101 et 103.
Le thésaurus pour l'écrit se compose généralement de deux ou trois grandes parties :

- Une liste alphabétique incluant les descripteurs avec leurs relations hiérarchiques, éventuellement des relations d'équivalence et des relations associatives, et les non-descripteurs
- Des champs sémantiques regroupant les descripteurs autour d'un thème commun
- Un index permuté permettant de regrouper les descripteurs par différentes entrées

Destiné à l'image, un thésaurus doit multiplier, en plus des relations hiérarchiques, les relations d'associations et d'équivalence entre les descripteurs ainsi que les notes d'application. Cette méthode permet effectivement de réduire les problèmes de polysémie et de relier entre eux des termes abstraits et concrets.
Il est souvent accompagné de listes additives telles que celles citées dans la catégorie « langage contrôlé » : liste morphologie, liste connotation, listes noms géographiques, …

Un thésaurus image peut être crée a priori, sans l'existence d'un vocabulaire d'indexation, ou a posteriori en prenant compte du vocabulaire d'indexation utilisé. La création d'un thésaurus a posteriori semble la plus adaptée pour le traitement des images, chaque source d'images possédant en effet des termes d'indexation qui lui sont spécifiques et qui sont difficilement applicables à d'autres fonds.

Dans le secteur de l'image les thésaurus sont généralement bilingues comme à l'AFP ou à France 2.

La plupart des logiciels spécialisés dans l'image proposent la matrice d'un thésaurus, comme nous le verrons dans la troisième partie (« Quelques logiciels spécialisés dans l'image »), mais nous avons remarqué que de nombreuses banques d'images fonctionnent sans.

ETUDE DE L'EXISTANT

I. L'agence de presse et d'illustrations CIRIC

1. Historique de l'agence

Ciric est une agence d'illustrations et de reportages spécialisée sur les questions sociales et les religions. Elle a été créée en Suisse dans les années 60 par des photographes soucieux de rassembler des témoignages visuels sur les chrétiens dans le monde et sur les enjeux du développement des pays pauvres. En 1981 l'agence a ouvert en France et s'est recentrée sur les questions sociales françaises : le développement, la santé, l'éducation, avec le soutien de plusieurs ONG. L'agence a élargi ses reportages à toutes les religions du monde (lieux de culte, rites, pèlerinages, art religieux). Elle a également multiplié en France et en Europe des portraits de personnalités de la vie religieuse. Son fonds est de 4 millions de photos (ektas et négatifs). Les photos numérisées disponibles sur la base et le site internet sont au nombre de 45 000.

En novembre 2003, Bayard Service Edition a repris les actifs de l'agence, qui était alors en liquidation judiciaire. Bayard est un groupe de presse, d'édition et de media, présent dans 41 pays sur les marchés de la presse jeunesse, des publications religieuses, de la presse senior. C'est le 5ème groupe de presse français par la diffusion. En 1992, ce groupe acquiert la SCPP, éditeur de presse paroissiale à Lille, et développe sa présence en régions sous le label Bayard Service Edition, qui assure un accompagnement régional des communautés chrétiennes (diocèses, paroisses, mouvements).

L'agence CIRIC est composée aujourd'hui d'une rédactrice en chef, d'une documentaliste, d'une assistante, de deux photographes salariés, d'une quinzaine de photographes réguliers et d'une trentaine de photographes occasionnels.
L'agence a décidé de moderniser sa base de données, jugée par ses clients archaïque, complexe et peu conviviale, dans le but d'améliorer ses ventes. Elle a décidé d'acquérir le logiciel Studio Orphea 3.6, déjà utilisé par plusieurs agences de photos (AKG, Rue des Archives, Gamma, …). Outre ses fonctions documentaires ce logiciel propose aussi un volant de gestion commerciale qui convient parfaitement à cette petite équipe. Par ailleurs Algoba proposait un transfert complet et sécurisé des données, transfert aléatoire pour certaines autres sociétés proposant un logiciel sensiblement similaire.

2. Typologie des usagers du CIRIC

Comme toutes les agences photographiques aujourd'hui, Ciric tente de tisser des liens étroits et à longs termes avec les rédactions de plusieurs titres de presse, en leur proposant des sujets sur mesure, hors des sentiers battus.

Le volume de vente de photos étant en perte de vitesse au Ciric, une enquête a été entreprise en février lors de rencontres ou d'entretiens téléphoniques avec des iconographes des journaux clients du Ciric :
Le Pèlerin, L'Express, Famille Chrétienne, Nathan, Phosphore, Croire Aujourd'hui, Le Figaro, La Croix.
Ces acheteurs ont renoncé dans leur majeure partie à faire leurs recherches iconographiques sur la base du Ciric car ils la trouvent trop complexe, trop lente, nécessitant « une trop grande gymnastique mentale ». C'est la recherche simple "contenant tous les mots" qui est plébiscitée massivement.
Les principaux griefs :

- temps d'affichage de la mosaïque et de l'agrandissement de la photo extrêmement lent
- définition pas assez précise pour les photos en haute définition
- ergonomie peu pratique (exemple : beaucoup trop de clicks pour accéder à l'information recherchée)
- absence de format IPTC (obligation d'imprimer la photo d'un coté, la légende de l'autre).

Notre enquête menée auprès d'iconographes a permis également de révéler que les fonctionnalités les plus appréciées sur les banques de photos en ligne sont :

- l'affichage des mots- clés qui ont servi à indexer la photo (Magnum)
- l' affichage des mots- clés associés à une recherche (Rue des Archives, AKG)
- le numéro de photo comme champ de recherche possible (Corbis)
- des dossiers thématiques (Réa)
- une indexation connotative (Corbis)

Selon ces utilisateurs la base fonctionne mal et le métier d'iconographe tel qu'il se pratique aujourd'hui ne laisse absolument pas « le temps d'errer sur une base mal faite » ! Ces personnes, conscientes de la spécificité du Ciric, se trouvent donc dans l'obligation d'y effectuer leurs recherches par téléphone.

3. La base de données CIRIC

La base de données du Ciric existe depuis 1983, avec une informatisation sous le logiciel Oracle (environnement DFI) depuis 1988. A partir de 1997 l'agence a confié également à DFI l'extension de sa base de données afin de développer la gestion d'images numérisées. Si DFI est parvenu à développer techniquement la base historique de l'agence pour la gestion des images à l'interne, elle a eu des difficultés à proposer une interface web client pour l'agence, n'arrivant pas à régler le problème des téléchargements et de la surveillance des connexions par l'agence.
Ciric a donc confié son interface web à un « hébergeur », la S.A.C.D (Société des Auteurs et Compositeurs Dramatiques). Cette dernière s'engageait alors dans une réflexion sur le développement de portail avec consultation

sécurisée des photos. Ce serveur permettait aussi le référencement internet de l'agence sans attendre le développement final de l'outil DFI. LA S.A.C.D proposait un produit non finalisé : le téléchargement des images était impossible, seule la consultation des images fonctionnait mais avec un moteur de recherche très rudimentaire (pas de croisement possible des mots-clés par exemple). En 1999 suite à un montage de dossier d'aides auprès du « Fond de modernisation des agences de presse », le Ministère des Finances a alloué une subvention au Ciric dans le cadre de son développement informatique. La Société DFI a donc retravaillé sur le projet de l'interface client. L'accès a été donné aux professionnels et clients du Ciric qui se connectaient par une adresse IP. Ceux-ci étaient bien souvent perdus entre ces deux interfaces proposées par l'Agence. Le produit DFI qui se présentait comme un projet performant en 1997 se retrouvait quasi obsolète dans le cadre de l'interface web : 5 ans après de nouveaux logiciels plus performants avaient inondé le marché. Suite à l'acquisition du système développé par Orphéa en août 2004, les informations de la base de données historique suivie des images numérisées liées aux fiches de reportages ont basculé dans un nouveau système de gestion ORACLE .

4. Le lexique de la base

La base gérait 45 000 photos numérisées et leurs informations textuelles.
En 2004 le lexique global de la base, regroupant tous les index (photographes, lieux, personnalités, mots- clés, …) comprenait **117 225** termes (aussi bien des mots que des chiffres correspondants au classement physique des photos). Les mots- clés étaient au nombre de **45 627** et correspondaient à des mots ayant servi à l'indexation ainsi qu'à des dates, des noms de photographes, des noms de personnalités, ect. Il y avait une redondance massive entre les termes de l'indexation (partie « mots-clés ») et les termes présents dans les différentes rubriques appelées « descripteurs ». Le lexique s'était accru ainsi de façon exponentielle et nous verrons les nombreux problèmes que cela a occasionnés et les solutions qui ont été étudiées pour l'améliorer.

La saisie de la fiche signalétique de la photo (ou du reportage entier) sur l'interface d'indexation de DFI se faisait sur une fiche documentaire comprenant les rubriques suivantes :

- n° reportage ou n° d'image si indexation par image (champ obligatoire)
- descriptif (= légende. Champ obligatoire)
- diffusion
- date (champ obligatoire)
- photographe (champ obligatoire)
- n° planches contact
- détail photos (=mots-clés)

Le n° de la photo (n° d'objet) était attribué automatiquement par la base de données.

La fiche de reportage avait 2 numéros : 1 numéro attribué automatiquement par la base de données, et un numéro correspondant au plan de classement thématique des diapositives dans la photothèque.

Exemple plan de classement :
5. SPORTS . LOISIRS . VIE ASSOCIATIVE
6. ENFANTS
7. FAMILLE . COUPLES
8. JEUNESSE
9. NOEL
10 . METIERS
11. DIVERS SUJETS DE SOCIETE

Les photographies étaient légendées par lots (souvent constitués d'une ou plusieurs planches contact) sur des fiches de reportages. Le texte de la fiche reportage était envoyée au client lors d'une commande en texte joint à la photo.

Les mots-clés étaient attribués par photos.

Les « descripteurs » (appelés ainsi dans la base car ce sont les rubriques sur lesquelles s'effectuaient les « recherches par descripteurs ») étaient les suivants :

- continent
- religion
- société
- vie quotidienne
- pays
- économie
- type de personne
- photo
- photographe
- n° reportage
- n° d'objet
- diffusion
- descriptif
- date

Certains de ces descripteurs possédaient une liste d'autorité. Cliquer sur l'une ou l'autre de ces zones permettait alors de visualiser son index et le nombre d'occurrence du terme dans la base.

Exemple :
RUBRIQUE « ECONOMIE »
agriculture : 564 occurences
artisanat : 164 occurences
atelier : 261 occurences
chantier : 201 occurences
commerce équitable : 73 occurences

L'indexation fonctionnait en langage libre, et reposait avant tout sur la mémoire des documentalistes chargés de l'indexation : pas d'outils d'aide à l'indexation.

Il n'y avait ni langage classificatoire ni thésaurus.

On notait, comme dans la plupart des banques d'images, une nette tendance à l'indexation pré-coordonnée :

Exemple :
FEMME
FEMME AGEE
FEMME AU FOYER
FEMME AU TRAVAIL
FEMME AVEC ENFANT
FEMME BATTUE
FEMME DE 60 ANS ET PLUS
FEMME DIACRE
FEMME EN PRIERE
FEMME ENCENINTE
FEMME EVEQUE
FEMME PASTEUR
FEMME POLICIER
FEMME PRETRE
FEMME RABBIN
FEMME SOLDAT
FEMME VOILEE

Il y avait très peu d'analyse connotative et très peu d'analyse morphologique.

Des « définitions » accompagnaient un grand nombre de mots-clés.

La recherche se faisait par l'image ou par la fiche textuelle.

A la recherche en interne on avait le choix entre :

- une recherche sur tout le lexique (avec beaucoup de bruit pour résultat, exemple :

En faisant une recherche en texte intégral avec le terme « oiseau » on trouve toutes les photos contenant le mot-clé « oiseau », mais aussi toutes les photos avec « oiseau » dans la légende mais ne comprenant pas d'oiseau dans les mots- clés : « Maire-Jeanne, 83 ans, vivant seule chez elle. Son quotidien est rythmé par la visite de ses enfants et par la compagnie de ses oiseaux. »

- une recherche par descripteurs
- une recherche sur les mots-clés.

Les opérateurs booléens fonctionnaient bien.

Mais on ne pouvait pas croiser plusieurs mots dans la rubrique mots-clés. [annexe 2]

Dans la recherche côté client (sur adresse IP : http//172.29.2.208) la recherche proposée était très simple : texte intégral ou par descripteurs. [annexe 3].

Il n'y avait aucune aide a la recherche. Il était impossible de vérifier que le mot tapé dans la rubrique « mot-clé » existait ou non dans la liste des mots-clés. Il n'y avait pas de croisement possible des mots- clés. Au résultat on constatait un bruit énorme.

II. le transfert de la base vers le nouveau logiciel choisi

Dans cette partie nous développerons essentiellement la méthodologie utilisée tout au long de notre travail. Nous donnerons quelques exemples concrets sur le travail de réflexion et de transformation entrepris.

1. Nettoyage du lexique

La première étape a consisté à récupérer la liste alphabétique du lexique global pour obtenir une liste alphabétique des termes utilisés et de leurs occurrences dans la banques de donnée. Ceci a nécessité l'importation par l'informaticien de Bayard, sous excel, des 2 fichiers provenant de la base Oracle et correspondants au « lexique » et aux « mots- clés ».
Le lexique comprenait 117 225 termes.

Suppression des fautes d'orthographes et de frappes

Un premier travail a consisté à nettoyer manuellement toutes les fautes d'orthographe, les fautes de frappe et autres scories ; ce premier travail a permis de supprimer tous les termes qui s'étaient accumulés au cours des années et qui n'avaient aucune utilité.

Suppression des occurrences nulles

Une évaluation statistique (élevée, moyenne ou faible occurrence) puis qualitative des termes utilisés a permis de supprimer beaucoup de mots. Une fois ce premier travail effectué il a fallu choisir, dans le but de proposer un vocabulaire unique, contrôlé et cohérent, les termes à conserver, ceux à supprimer (obsolescence, doublon, …) et ceux à remplacer.

Exemple de termes à supprimer :
AVIS DE DECES
ANTI CRIQUETS
ANTI EROSION
ANTI EXCLUSION
ANTI VOL
ARONDE
APOTHICAIRE
ALIMENTATION DU BETAIL
CENTRE D'AVEUGLES
CERTIFICAT
CENTRE
CERDON
CEREMONIAIRE
CERNE

Exemple de termes a remplacer :

BUS	AUTOBUS
BUS SCOLAIRE	RAMASSAGE SCOLAIRE
BUT	FOOTBALL
BEAU-PERE ET ENFANT	BEAU-PERE AVEC ENFANT
BINAGE	JARDINAGE
BOUCHE DE METRO	METRO
BALADE	PROMENADE
BELMONT	ABBAYE DE BELMONT
BYZANTINE	ART BYZANTIN
B.R.E.D.	BANQUE
BILLET DE BANQUE	ARGENT
CHRETIEN CELESTE	SECTE PROTESTANTE

Ce nettoyage a permis de supprimer du lexique global de la base plus de **83 000** termes.

2. La prise en main du nouveau logiciel

Le logiciel Studio Orphea 3.6 a été conçu par la société Algoba, spécialisée dans le développement de logiciels de gestion d'informations électroniques dédiées aux documents multimédias. Ce logiciel bénéficie d'une certaine reconnaissance sur le marché des banques d'images tant françaises : Sipa Press, REA, Gamma Presse Images, AKG, Godong, Œil Public, Tendance Floue… - qu'internationales : Reporters Press Agency (BE), Redux (USA), La Presse (IT), TiPress (CH), Splashnews (USA, UK), …

Organisé autour de la base de données Oracle, il permet de cataloguer de manière automatique ou manuelle dans une seule et même base des documents de tous types : images, Adobe PDF, video et son.

Ses principales fonctionnalités pour la partie images sont les suivantes :

- Administration : Gestion des utilisateurs et de leur restrictions (accès sécurisé par adresse Ip de l'utilisateur, par mot de passe). Création de reportages, application de watermarks ... statistiques sur les connexions, les recherches, les consultations, les téléchargements.

- Gestion assistée du langage documentaire : Indexation par mots-clés et indexation en texte intégral pour les champs titre et légende . Extraction des champs IPTC / JPEG des images référencées (si le média concerné est de la photo en format JPEG, TIFF ou EPS). zone IPTC en format unilingue, dans la langue pivot de la banque d'images. Lien IPTC/ lexique et thésaurus.

- Affichage des résultats : Ergonomie entièrement paramétrable . affichage du résultat sous forme de en mosaïque (3 x 2 , 3 x 3, 4 x 3 ou 5 x 4) ou de listes (de 20, 40 ou 60 lignes). Accès au zoom. lien vers les autres images du reportage. Fonctions de retouche avant archivage ainsi qu'un module d'annotation et de cadrage.

- Gestion du workflow : circulation logique et physique des documents dans un groupe de travail. Envoi automatique d'un document par e-mail et/ou par FTP. Gestion commerciale, facturation en batch et par lots, calcul des droits, suivi des clients et de l'historique de leur activité.

- Multilinguisme
  6 langues. dictionnaire de 100 000 mots français et 26 000 mots anglais

Conséquences sur l'indexation pour la base de données du Ciric

L'agence a repensé sa politique d'indexation qui jusqu'à présent privilégiait la surindexation et négligeait totalement l'aspect connotatif.
L'affichage systématique des mots-clés lors de l'indexation, à côté de la photo elle même, va lui apporter une aide considérable ; la zone mots-clés reste un index ouvert et des termes « candidats » peuvent être validés.
Le traitement des images ayant un tronc commun au niveau de l'indexation va être allégée puisqu'on peut dupliquer les mots-clés et les appliquer au nombre de photos choisies.
A l'échelle de toute la base on peut aussi remplacer un mot-clé par un autre et appliquer ce changement à toutes les fiches retenues, ce qui facilitera le changement de certains mos clé devenus obsolètes.
La partie thésaurus est jugée très intéressante par l'agence (surtout pour les synonymes , les notes d'application) qui espère la développer plus tard.

Changements pour la recherche

Le logiciel Orphéa fonctionnant sous Oracle, beaucoup de fonctionnalités sont les mêmes. On retiendra donc ici seulement les différences :

- Les 3 modes de recherche (texte intégral, par rubriques, par mots-clés) permettent tous les trois de choisir entre plusieurs collections à l'intérieur du CIRIC, ce qui peut être utile dans le cas de restrictions de droit d'utilisation à certains clients.
- La recherche par descripteurs se fait sur les rubriques qui ont été au préalable retenues par l'administrateur de la base avec changement possible à tout moment, ce qui est plus souple que dans l'interface dfi (en général : légende, mots- clé, référence, photographe, titre). Par contre ce mode de recherche peut combiner 3 descripteurs maximum alors qu'auparavant les combinaisons de descripteurs n'étaient pas limitées.

- Une recherche sur une partie du mot est maintenant possible grâce aux troncatures.
- La recherche par mots- clés permet d'afficher sur l'interface web les mots-clés existants. Cela représente une véritable aide à la recherche, aide qui était totalement absente sur l'ancien site.

Exemple :
Si je tape le mot ENFANT le moteur de recherche me proposera avant de lancer la recherche :
ENFANT
ENFANT A LA FENETRE
ENFANT ABANDONNE
ENFANT AU TRAVAIL
ENFANT AVEC POUPEE
ENFANT BOUDANT
ENFANT DANS L'EAU
ENFANT DANS LES BRAS
ENFANT REFUGIE
ENFANT SOLDAT
ENFANT DES RUES
Ect
En me précisant le nombre de photos correspondants aux mots- clés. Ce n'est qu'ensuite que je choisirai mon mot-clé.

Cette recherche permet de sélectionner plusieurs mots-clés mais uniquement ceux qui se suivent, par conséquent la recherche sur plusieurs mots-clés n'est toujours pas vraiment possible, ni pour l'administrateur ni pour le client.

Exemple :
Je pourrais activer la recherche sur ENFANT AU TRAVAIL + ENFANT AVEC POUPEE , mais pas ENFANT AU TRAVAIL et ENFANT DES RUES. Je ne pourrais pas davantage cliquer sur ENFANT SOLDAT avec REFUGIE.

Par contre l'activation de la rubrique « mots- clés associés » est pertinente ; elle permet par exemple pour une photo affichée avec le mot-clé :
CAMP DE REFUGIES
de pointer les mots- clés associés à cette image (qui ont un lien hypertexte et qui relancent une autre recherche sur un simple click) :

COLONIES DE REFUGIES
ENFANTS REFUGIES
HAUT COMMISSARIAT AUX REFUGIES
REFUGIES POLITIQUES

Le véritable changement opéré par le nouveau logiciel me semble être dans aide en ligne apportée aux clients.

3. Applications de nouvelles règles d'écriture

Quelques nouvelles règles d'écriture ont été proposées à l'équipe et acceptées :

Majuscule
Tous les mots- clés sont systématiquement écrits en majuscule pour une homogénéité visuelle.

Genre et Nombre
Le choix de garder uniquement les singuliers a été validé dans la mesure du possible.

Noms propres
Ils sont de différents types : noms de personnes, de termes géographiques,, d'organismes …
Pour l'ordre des noms de personnes il a été retenu :
Le « nom » suivi du « prénom »
Quand c'est un portrait on a précisé :
« Nom prénom (portrait) »

Particules
Elles ont été conservées en début de nom.
exemple : DE MAZENOD EUGENE

Suppression des points
Après réflexion les points ont finalement été maintenus pour ne pas créer du silence lors de recherches futures. Exemple : S.O.S Racisme n'apparaîtra pas si on le cherche avec SOS Racisme donc on maintient son écriture initiale.

Dates
Les dates sont écrites en : jours en chiffres / mois en toute lettres / année en chiffres.
Exemple : 15 avril 2004

Sigles et abréviations
Il a été choisi de conserver les sigles et d'inscrire en plus leurs développés lorsque cela n'était pas déjà fait dans la légende, en vue de la recherche en full text ou par descripteurs.
exemple : HCR (Haut Commissariat aux Réfugiés)

En revanche pour les mots- clés il a été maintenu l'appellation par sigles dans la mesure où celle-ci est la plus connue.

Exemple :
O.N.G figure parmi les mots- clés mais pas ORGANISATION NON GOUVERNEMENTALE.

Nationalités
Une réflexion a été engagée pour éviter à la fois la redondance à l'indexation et les confusions de résultat.

Exemple :
Dans l'ancienne base la photo d'un homme mexicain était indexée à : « HOMME MEXICAIN » tandis qu'une femme était indexée à « FEMME MEXICAINE ». Cette pratique n'a pas été modifiée alors qu'il a été envisagé d'indexer ce type de photos à « HOMME » ou « FEMME » ET « NATIONALITE MEXICAINE » pour éviter les confusions à la recherche. Mais cette décision n'a pas été prise faute de temps.

4. Améliorations du langage documentaire

Les définitions

Des définitions avaient été intégrées dans la base au fil des années mais n'étaient plus toutes valides, d'une part en fonction d'une non-pérennité des informations, d'autre part en fonction du peu d'intérêt qu'elles suscitaient (ex : agouti ou albinos).
Par contre en fonction de la spécificité du fonds il a été jugé opportun d'en conserver certaines, même en les modifiant, dans le but de les importer sur la nouvelle base.

exemple :
ALASITAS: figurines utilisées comme amulettes dans les cultes syncrétiques des pays andins.
ALTERMONDIALISTE: partisan d'une forme d'organisation des rapports mondiaux plus favorables aux pays pauvres.
AMANTE DE LA CROIX: congrégation religieuse.
AMBON: pupitre ou petite chaire à partir d'où se font les lectures et l'homélie de la messe.
AMBROSINIEN: groupement de laïcs attaché à la tradition cistercienne autour de Citeaux et des lieux de fondation de l'ordre.
ANCOLI: rassemblement de chorales religieuses catholiques

Cette réflexion a été menée avec la consultation du dictionnaire des religions et des mouvements philosophiques associés disponible sur le world wide web : http://atheisme.free.fr/Religion/Lexique.htm

Les synonymes

Une proposition de synonymes a été faite également suite à des réunions avec les deux personnes chargées de l'indexation.

Exemple :

OBSEQUES	FUNERAILLES
ENTERREMENT	OBSEQUES
FUNERAILLES	ENTERREMENT
PRECARITE	PAUVRETE
GENS DU VOYAGE	TZIGANES
ROMS	TZIGANES
ROMS	GENS DU VOYAGE
DECES	MORT
DEMON	DIABLE
CURE D'ARS	VIANNEY JEAN BAPTISTE
FIDEI DONUM	PRETRE

Le thésaurus

Une réflexion a été menée pour déterminer les besoins de l'agence par rapport à un thésaurus, dans la mesure où le logiciel acquis en proposait une structure. Dans ce but une recherche sur internet a été menée pour trouver un travail initial déjà réalisé dans le domaine des religions.
Le « thésaurus international de la terminologie relative aux réfugiés », réalisé dans le cadre des Nations-Unies en 1996, comporte un champ sémantique (MICRO-THESAURUS 12) sur l' appartenance religieuse. Il s'agit d'une liste hiérarchique à trois niveaux qui aurait pu être intéressante comme première approche de ce thème (site : http://www.huridocs.org/mtf.htm).

Mais l'agence s'est rendue compte qu'elle n'avait actuellement pas les moyens humains ni temporels pour se lancer dans un travail de cet ordre et cette recherche a été interrompue

Propositions de listes d'autorité

Plusieurs listes d'autorité ont été proposées pour palier les manques constatés.

- POUR LES PERSONNES

Enfant de 0 à 3 ans
Enfant de 3 à 13 ans
Femme
Fille
Garçon
Homme
Jeune de 13 à 18 ans
Personne de 18 à 35 ans
Personne de 35 à 60 ans
Personne de 60 ans et plus

- POUR LES TYPES HUMAINS

A utiliser dans un contexte de description de minorité

type africain
type européen
type indien d'Amérique
type indien d' Asie
type asiatique
type arabe
type latino-américain
groupe multiracial

- POUR LA TYPLOGIE IMAGE

Contraste
Contre-jour
Contre-plongée – synonyme : vue d'en bas
Couleur
Détail
Extérieur
Flou
Gros plan – synonyme : plan rapproché
Intérieur
Lumière
Matière
Mouvement - synonymes : bougé, filé
Ombre
Plongée – synonyme : vue de haut
Portrait
Portrait
Vue aérienne
Vue circulaire
Vue de dos
Vue d'ensemble - synonyme : vue générale
Vue de loin
Vue panoramique

III. Application de la méthode du Benchmarking pour la construction de l'interface web

Avant la mise en ligne de la base de données du Ciric sur la nouvelle interface web proposée par le logiciel Studio Orphea 3.6, un travail de benchmarking a été réalisé afin de déterminer les choix de l'agence en matière de présentation de sa base.

Le benchmarking est un outil qui permet de se doter de moyens pour progresser et conquérir de nouveaux clients. Cette pratique oblige à aller chercher à l'extérieur les meilleures méthodes employées afin de les incorporer à son propre fonctionnement.
Le benchmarking permet de se poser les bonne questions : comment font les autres ? Peut-on faire aussi bien ou mieux ?

Eric Suttter propose dans son ouvrage 4 axes de prospection [32] :

- Le benchmarking interne
- Le benchmarking auprès des concurrents directs
- Le benchmarking orienté vers d'autres fonctions (sortant de son univers habituel et de son secteur d'activité)
- Le benchmarking horizontal (certains processus demeurent identiques quelles que soient les disparités entre les activités)

Dans le cadre du Ciric c'est le benchmarking auprès des concurrents directs qui a été privilégié.

Les différentes phases de benchmarking ont été les suivantes :

- identifier l'objet des recherches
- identifier les sites à comparer
- effectuer la comparaison
- communiquer les résultats du benchmarking et les faire accepter
- démarrer les actions correspondantes

Objet des recherches
L'interface web de l'agence Ciric était jugée jusqu'ici inopérante pour tous les problèmes que nous avons déjà listés. Les améliorations à apporter en terme de présentation et d'ergonomie étaient nombreuses.

Après avoir repertorié les défauts de l'ancienne interface, nous avons étudié les fonctionnalités qui allaient être amélioré avec le nouveau logiciel :

- la mise en avant des derniers reportages faits par l'agence
- la création de portfolios et de sélections

- la téléchargement à l'unité et par lots
- la commande en ligne
- la création de paniers enregistrés
- les paramétrages pour l'interrogation et la recherche

Comparaison

Toutes ces fonctionnalités web existent, avec quelques variantes, sur les sites des agences photographiques. Nous nous sommes concentrés sur les qualités de présentation et d'ergonomie d'une vingtaine d'interfaces web d'agences. Pour certaines agences il a été nécessaire de prendre contact et de faire des demandes d'accès privé et autonome (mot de passe et code identifiant). Ces codes m'ont été accordés en fonction de mon statut d'étudiant et ont fait l'objet d'un accord de non-usage commercial de ces bases.

Voici la liste des url des bases photographiques étudiées :

http://www.afp.com/francais/products/?pid=image/imageforum

http://www.agencevu.com

http://www.akg-images.com

http://www.bsip.com/fr/

http://pro.corbis.com

http://www.editing.fr

http://www.gamma.fr

http://creative.gettyimages.com

http://www.godong.fr

http://www.lumieredumonde.com

http://www.magnumphotos.com

http://www.oeilpublic.com

http://www.reaphoto.com

http://www.ruedesarchives.com

http://www.rapho.com

http://www.roger-viollet.fr/principale.asp

http://www.sipa.com

http://www.urbaimages.fr/wimsolurba/

Beaucoup d'agences ont des présentations simples, ergonomiques, ne nécessitant aucune compétence informatique documentaire. L'impression générale est souvent bonne. La navigation est souvent simple, proposant des barres de navigations en haut de page, des boutons pour se déplacer,

mais peu de plans de site. Sur certains sites (principalement les portails) il est en revanche difficile de répertorier les différentes collections et l'on ne sait pas toujours dans quelle(s) base(s) la recherche demandée s'effectue (cf. Corbis). Dans d'autres la profusion d'images proposées est telle qu'il est difficile de se repérer (cf. Magnum).
La densité informationnelle est acceptable et l'implantation des zones bien organisées (sauf sur des sites où l'information est linéaire et nécessite des ascenseurs sur chaque page, comme Imageforum ou Rapho),
Les sites sont d'aspect graphique, en phase avec leur contenu photographique.

En ce qui concerne le design , on note une certaine homogénéité dans les couleurs dominantes :

- du gris (Rapho , Magnum, Rea, Editing, Urba images, …)
- du blanc (Akg, Bsip, Getty, Hachette Filipacchi Photos, agence Vu, Roger Viollet, …)
- du noir (Contact press images, Godong, Image Forum, l'Oeil public, Rue des archives,…)

Le mode positif est le plus souvent favorisé (fond clair avec caractères foncés). Le nombre de couleurs utilisées est limité (3 voire 4 au maximum).

Beaucoup d'agences ne permettent pas d'entrer sur leur site sans mot de passe , la page d'accueil étant alors limitée à de 2 ou 3 petites photos, une photo centrale plus grande, et une fenêtre d'enregistrement de son code d'accès. C'est le cas des agences Godong, L' AFP, Roger Viollet , Gamma et AKG.

D'autres en revanche proposent un très grand nombre de photos, protégées par des watermark, sans mot de passe, ordonnancées dans des rubriques bien identifiables (cf. Lumiere du monde, Rapho, Rue des archives , ANA , Corbis) :

- Les reportages ou nouveaux reportages
- Les actualités (encore qu'il soit souvent difficile de déterminer les différences entre les deux)
- Les Top sélections
- les portfolios des photographes
- les expositions

La traçabilité du parcours n'existe pas toujours, ni même l'historique de la recherche. Il arrive très fréquemment que l'oon se perde dans le dédale des navigations faisant passer l'utilisateur d'un reportage à un autre sans toujours le lui préciser.

<u>Choix établis pour Ciric</u>
Pour le design, le choix a été pour un fond gris clair, jugé très opérant pour faire ressortir les photos, assorti de blanc et de rouge. [annexe n°4]

Pour l'accès, le choix du Ciric a été de ne rien proposer sans loggin. Seules 3 photos sont visibles, dont le remplacement a été prévu toutes les semaines. La page d'accueil propose également la présentation de l'agence et le contact (il a été convenu que sa page d'accueil mettrait en avant sa spécificité : l'information religieuse, sociale et humanitaire et le portrait de personnalités)

Sur la page suivante il a été prévu de présenter systématiquement les « nouveaux reportages » (une quinzaine), avec accès dans la barre de navigation aux « actualités » et dans la frame de gauche aux « top sélections », avec une mise à jour de toutes ces sélections environ tous les mois. [annexe 5]. L'avantage de cette solution est une véritable dynamique n'imposant pas de mise a jour trop fréquente, ce qui aurait été une trop lourde contrainte pour l'agence.

En ce qui concerne les portfolios des photographes il a été convenu de construire cette page d'après l'exemple de l'agence Vu : une liste de photographes sur la gauche (avec 2 parties : les membres et les distribués) et un lien vers leurs reportages , tout en gardant l'éventualité de rajouter une rubrique « le photographe du mois ».

Aucune agence visitée, même celles ayant un fonds très spécifique, ne proposent de définitions en ligne. L'ancienne base de données du Ciric en interne en possède beaucoup, et ceci a paru très intéressant pour la nouvelle mise en ligne, car la base regorge de mots- clés pas toujours compréhensibles pour les clients. Mais si l'importation des définitions est aujourd'hui possible sur le logiciel d'Algoba en interne, la mise en ligne de ces définitions sur l'interface web n'est pas encore réalisable et nécessiterait une nouvelle version du logiciel.

Toutes les agences visitées ou presque proposent une aide en ligne pour expliciter les différents modes de recherche, la création d'un compte, l'affichage des résultats, la sauvegarder un panier, le téléchargement, la personnalisation de ses préférences…
L'aide conçue par Algoba étant efficace, celle-ci a été conservée.

ETUDE COMPARATIVE DU TRAITEMENT DOCUMENTAIRE DANS LES AGENCES PHOTOGRAPHIQUES EN LIGNE

I. Les fonds d'images en France

Selon le répertoire Iconos de la Documentation française de 2004, on recense aujourd'hui en France :

- environ 160 agences photographiques, parmi lesquelles :

- des agences de presse télégraphiques (Agence France Presse, Reuters, Associated Press, United press International).
Leur politique de production est axée sur une couverture quotidienne et mondiale de l'actualité. Elles diffusent par abonnement des dépêches écrites et des images. AP a crée son service photo en 1935, UPI en 1958, l'AFP en 1958 et Reuters en 1985. Outre la diffusion systématique à leurs abonnés elles offrent aussi un accès à leurs archives via internet. L'AFP a ouvert son service en ligne en 1999.

- des agences d'actualité et de news (Gamma, Sygma, Sipa, toutes rachetées par de grands groupes aujourd'hui…).
Ces entreprises commerciales produisent, diffusent, conservent et commercialisent des reportages photographiques. Elles couvrent l'actualité immédiate mais proposent aussi des reportages approfondis autour de divers sujets (« features ») ainsi que des portraits ou reportages « people », en accroissement aujourd'hui. Etant donné le volume énorme de leur fonds, ces agences ont rapidement dû s'informatiser. Chacune a également étendu sa politique commerciale en rachetant de plus petits fonds. Mais le rachat par la société Corbis de l'agence Sygma en 1999 a marqué le tournant dans l'histoire des agences photographiques françaises comme nous le verrons plus loin.

- des agences de photographes (Magnum, Vu, Editing, Contact Press Images, Métis Images).
Elles ont été créées par des photographes indépendants soucieux de préserver leur autonomie et de contrôler l'utilisation de leurs images par la presse. On assiste depuis quelques temps à la création de nouvelles petites agences indépendantes regroupant des collectifs de photographes animés du même esprit : Tendance Floue, L'œil public, Tango Photo, Dolce Vita, agence VII … Certains photographes gèrent seuls leurs fonds. C'est le cas de Sebastiao Salgado avec Amazonas Images Presse, de Martine Franck, de l'atelier Charmet, ect…

- des agences d'illustration (Giraudon, Image Bank, rachetées par Getty Images, Hoa Qui, Explorer, Jacana, rachetées par Hafimage,…).
Leur organisation est différente . Les images proviennent de reportages aussi bien que de prises de vue réalisées sur commande en studio avec des figurants ou des mannequins.

- des petites agences spécialisées dont beaucoup ont été rachetées par les groupes Getty, Corbis et Hafimage :
en architecture, comme Archi Press ou Urba images
en histoire de l'art – AKG photo, Edimedia, Artephot
en sciences - Ciel et Espace, Cosmos
en théatre – Enguerand, Bernand
…

Les fonds d'images français se composent également de :

- près de 700 photothèques d'organismes publics, musées, bibliothèques, archives …

La bibliothèque Nationale détient la plus importante collection au monde, répartie sur plusieurs de ses départements (Estampes et photographie, Cartes et Plans, Musique, Manuscrits,) grâce à l'instauration du dépôt légal. Une sélection est consultable en ligne (www.gallica.fr). De plus en plus de musées détiennent des collections d'images organisées autour de thèmes. Les administrations et collectivités locales possèdent aussi des fonds importants (CNRS, Ministère de l'Agriculture, des Transports, de la Défense) ou les développent dans les secteurs communication des mairies, conseil Généraux ou conseils régionaux.
Les établissement publics comme la RATP, le Gaz de France, la SNCF, gèrent ou font gérer leurs grands fonds d'images.

- 75 photothèques d'entreprises ou d'associations privées (Rhône-Poulenc, Elf Aquitaine, Michelin ….) Leurs images proviennent généralement de reportages réalisés pour l'illustration de leurs supports d'information, mais certaines possèdent de véritables fonds historiques.

II. L'évolution du marché de la photographie de presse et d'illustrations

Si la révolution numérique a bouleversé les techniques photographiques, elle a aussi transformé radicalement l'ensemble du commerce de l'image. La course frénétique à la concentration des sources d'images est aujourd'hui planétaire. Trois groupes se disputent le marché mondial de l'image :

- La société Corbis, propriété personnelle de M. Bill Gates, fondateur de Microsoft
- l'agence Getty Images, fondée par l'un des descendants d'un magnat du pétrole américain
- Hachette Filipacchi Medias (HFM), filiale presse du groupe français Lagardère, qui édite plus de 200 magazines dans le monde

Crée en 1989, Corbis a pris le contrôle de plusieurs agences de presse : Saba (news), Kipa (télévision et cinéma), Tempsport (sport) et Outline (people) et surtout Sygma (peut-être la plus grande agence photographique du monde), ainsi que de différentes agences de photos d'illustration et de collections comme le fonds Bettman (la plus grande mémoire photographique des Etats-Unis, 16 millions d'images) ou le Stock Market. Corbis a également signé des accords de reproduction avec la National Gallery et le musée de l'Ermitage. Elle possède aujourd'hui une collection de 68 millions d'images.

Getty Image s'est développé sur le marché de l'illustration avec comme clients des magazines, des maisons d'éditions et des agences de publicité. Elle a racheté les agences Tony Stone Images (illustration), Liaison (news), The Image Bank (la banque d'image de Kodak, agence d' illustration), PhotoDisc, Eyewire et Giraudon . Elle souhaite également mettre un pied sur le marché de la photographie d'actualité comme le montre son alliance avec l'AFP en 2003, bien qu'aujourd'hui encore 90 % de ses ventes soient des photos publicitaires. La société dispose aujourd'hui d'un catalogue de 70 millions d'images.

Hachette Filipacchi Medias, déjà propriétaire de l'agence généraliste Gamma (news) depuis 1999 , a poursuivi intensivement ses acquisitions en rachetant trois autres agences françaises spécialisées : Jacana (faune et flore), Hoa-Qui (tourisme, art de vivre) et MPA. HFM a aussi pris le contrôle en 2000 de l'agence Rapho (reportage et archives) et Keystone en 2001 (15 millions de clichés dont le fonds de l'Humanité jusqu'à 1970 et le fonds de l' Illustration). Forte d'un fonds de 40 millions de clichés, HFM affirme vouloir créer un « pôle européen de la photographie » et prépare la mise en ligne de son portail baptisé Hachette Filipacchi Photos, une filiale pour la vente en ligne de photos numérisées, provenant des agences acquises et de la production

des publications du groupe comme les hebdomadaires Paris Match (8 millions de clichés) ou Elle.

Parallèlement on assiste à un développement spectaculaire des services photos des trois principales agences de presse mondiales, Associated Press (Etats-Unis), Reuters (Grande-Bretagne) et Agence France Presse. Ce sont aujourd'hui principalement leurs images d'actualité qui illustrent les magazines. Mais elles non plus n'échappent pas au climat de concentration actuel . L'AFP s'est alliée avec Getty Images en 2004 et commercialise auprès de ses abonnés dans le monde, à l'exception des Etats Unis et du Royaume Uni, les photos de la couverture nord-américaine de Getty. Le portail ImageForum regroupe entre autres les bases de données photos de l'Afp, Getty, Roger Viollet, Photos12.

On voit aussi apparaître de nombreux collectifs explorant la voie de la vente directe par le biais de sites Internet dont les auteurs, de plus en plus conscients de leurs droits, se mobilisent très fortement au niveau mondial (l'Oeil Public, ExMundo, Transit, Tendance Floue, Tango Photo, Dolce Vita, agence VII…) Ce sont des associations ou des coopératives de photographes indépendants, qui essaient de proposer une alternative aux agences classiques. Mais réussiront-elles à s'imposer dans le monde de plus en plus étroit du photo-journalisme ?

Il faut aussi compter avec la montée en puissance de nouveaux acteurs qui se positionnent sur l'offre de clichés « libres de tous droits » et constituent de gigantesques bases de données de photos d'illustration. Des sociétés comme Librededroits.com, Présentation Direct, Diaf-SDP, vendent (parfois en passant des accords avec les groupes Corbis ou Getty) aux professionnels de la publicité, des médias, de l'édition, du multimédia et de l'Internet des photographies « libres de droits » téléchargeables à l'unité ou achetables par lots sur support CD Rom. Cette nouvelle forme de vente, encore à ses débuts en France, paraît en pleine expansion sur le marché européen et représente déjà 40% du marché aux Etats-Unis.

Sur fond de course effrénée à la croissance, ces mutations radicales s'accompagnent aussi de conflits en matière de droits d'auteur. Les photographes se sont très fortement mobilisés en France, s'élevant contre une « américanisation des contrats » proposés par leurs nouveaux interlocuteurs, qu'il s'agisse de la société Corbis ou de HFM. En effet les dispositions très protectrices du droit d'auteur français permettent à un photographe de décider du cadre dans lequel ses travaux pourront être exploités, en lui garantissant une rémunération à chaque nouvelle utilisation d'un cliché. Les photographes s'élèvent contre la tentative d'imposition de nouvelles règles, leurs clients s'arrogeant tous les droits d'exploitation et de commercialisation de leurs photos, moyennant une rémunération forfaitaire initiale, acquise une fois pour toutes. Ces conflits sont de plus en plus fréquemment portés devant les tribunaux et s'inscrivent dans un contexte de « juridicisation » croissante du droit à l'image, qui limite le travail

« documentaire » des photographes au profit d'un développement massif de l'image d'illustration, mettant également en péril la notion de photo-journalisme.

III. Quelques logiciels spécialisés dans l'image

1. Principales fonctionnalités des logiciels

Les orientations en matière de langages documentaires de ces banques d'images en ligne sont fortement tributaires des logiciels. Ceux-ci sont en pleine évolution depuis une quinzaine d'années. La puissance des matériels informatiques et des logiciels ainsi qu'une importante baisse des coûts a permis aux photothèques même de moyenne importance de s'équiper.

Deux types de solution existent aujourd'hui sur le marché professionnel :

- les logiciels documentaires ou SGBD (système de gestion de banques de données) gérant une base de données à laquelle est relié un imageur d'images.
- les logiciels intégrés proposant de la gestion multimédia (image fixe depuis plusieurs années et maintenant video et son)

Nous traiterons ici des seconds car ceux sont ceux les plus répandus dans le marché des agences photos.

Ces logiciels de gestion multimédia proposent tous un module de gestion documentaire et un module de gestion des images, avec parfois un module de gestion commerciale à acquérir en plus (cette option est beaucoup plus accessible financièrement qu'il y a une dizaine d'années).

Les fonctionnalités principales qu'on retrouve quasiment sur tous ces logiciels sont les suivantes :

- Gestion documentaire

Création, structuration, paramétrage et administration de la base.
Importation de données d'anciennes bases
Saisie et mises à jour des données
Importation des formats IPTC
Indexation des données
Gestion de thesaurus, de lexiques
Multilinguisme
Recherche (texte intégral, par descripteurs selon plusieurs critères définis par l'administrateur, par mots- clés, avec opérateurs booléens)
Gestion des statistiques, de la confidentialité

- Gestion des images

Acquisition : numérisation, compression, retouche, importation d'images, conversion de formats
Consultation des images à partir des notices documentaires

Gestion de l'affichage des images (vignette, plein écran , haute définition) et des textes
Exportation vers des CD Rom
Paramétrage de l'affichage, de l'édition et des exports

2. Grille d'évaluation de 5 logiciels

Nous verrons dans la grille d'évaluation suivante les quelques disparités existantes parmi les 5 logiciels professionnels les plus utilisés pour l'image fixe.

SOCIETE	ARMADILLO	ALGOBA	WEST VALLEY (groupe CITE de L'IMAGE)
site	http://www.armadillo.fr/	http://www.orphea.com/index_fr.html	http://www.web-valley.com/
LOGICIEL ETUDIE	ARMADILLO PHOTO MEDIA	ORPHEA STUDIO 3	WESTPHOTO
INDEXATION	Fiche descriptive standard de 14 rubriques. 7 rubriques contrôlées avec des listes. Processus de validations successives des candidats descripteurs. Aides à la saisie. Intégration des champs IPTC et EXIF. Indexation de tous les mots de toutes les rubriques. Intégration de thésaurus et de listes d'index . Gestion d'historiques et de statistiques	Fiche descriptive standard de rubriques. Processus de validations successives des candidats descripteurs. Aides à la saisie. Intégration des champs IPTC et EXIF. Indexation de tous les mots de toutes les rubriques. Intégration de thésaurus et de listes d'index. Gestion d'historiques et de statistiques.	Gestion de 1 ou 2 niveaux de catalogage (Reportage / Photos). Gestion de plus de 50 formats d'images standards : JOEG, TIF, GIF, EPS, PDS
RECHERCHE	Recherche en texte intégral, par descripteurs et par mots-clés. Opérateurs : et, ou, sauf, égal, entre, >, >=...	Recherche en texte intégral, par descripteurs et par mots-clés. Affichage des mots clés qui ont servi à l'indexation et des mots clés associés.	Recherche en texte intégral seulement ; l'utilisateur peut choisir dans la liste proposée les termes approchant.
GESTION COMMERCIALE	Envoi des documents par mails. Impressions sous de multiples formats, Gestion des droits d'auteurs. Gestion des commandes, des téléchargements et des prêts. Gestion du watermark. Diffusion sélective et gestion des droits de diffusion	Envoi automatique de document par email et/ou par FTP. Gestion des droits d'auteurs.Gestion des commandes, des téléchargements et des prêts. Facturation en batch et par lots.Suivi de clients et de l'historique de leur activité. Gestion du watermark. Système de paiement en ligne	Gestion des droits d'auteurs. Gestion des commandes, des droits d'accès, des téléchargements et des prêts. Gestion du watermark.
APPLICATIONS EN FRANCE	Tous secteurs dont : Vivendi Universal Education, RATP, ECPA-D , Château de Versailles, groupe Figaro, Réunion des Musées Nationaux,...,	Tous secteurs dont agences photos : Sipa Press, Gamma, AKG,), Godong, Œil Public, Tendance Floue...	Tous secteurs dont : CNRS Audiovisuel , Médecins Sans Frontières, Museum National d'Histoire Naturelle, Ciel & Espace
LANGUES	Gestion du multilinguisme Intégration de dictionnaires et de thesaurus	Disponible en Anglais, Français, Allemand et Italien. L'interface Web est disponible également en plus en Espagnol et en Arabe. Intégration de dictionnaires et de thesaurus	Française
OPTIONS	Hébergement, externalisation. Fabrication de CD Rom avec moteur de recherche embarqué. Création de sites web organisés comme des albums. Une photothèque internet réunissant des	Versions à plusieurs niveaux de prix (installation complète, location hébergement ou solution d'étagère) Module complet de « reporting » et de statistiques basé sur Business Object.	Conception de sites web et personnalisation graphique de l'application web

SOCIETE	ORKIS	GESCO (partenaire des sociétés CADIC et EVER)
site	http://www.orkis.com/Fr_Default.htm	http://www.gesco.fr/fr/Index.htm
LOGICIEL ETUDIE	AJARIS	IMAGEUR GESCO MEDIA
INDEXATION	Fiche descriptive de 35 rubriques Gestion de la norme IPTC dans les images au formats JPEG, TIFF et PSD. Intégration de thésaurus et de listes d'index. Les mots contenus dans un champ de texte libre peuvent être indexés à la demande. Modification de mots clé sur un lot de documents.	Intégration de thésaurus et de listes d'index. Aides à la saisie
RECHERCHE	recherche en texte intégral et par descripteurs. Recherches mémorisées.	recherche par rubriques et par mots-clé. Opérateurs : et, ou, soit, sauf.
GESTION COMMERCIALE	Gestion des commandes, des téléchargements et des prêts. Gestion du watermark.	Impression Exportation
APPLICATIONS EN FRANCE	Tous secteurs dont : groupe Pinault-Printemps-Redoute, agences de presse ou d'illustration : Top, Rapho, Abacapress, Keystone, Hoa Qui, Iconos. Gallimard, Paris Première,Terres d'Aventures	Tous secteurs dont : SNCF , Monuments historiques, Institut du Monde Arabe, Peugeot Citroen,conseils généraux, France Telecom ...
LANGUES	Jusqu'à 5 langues avec l'option Gold	Française
OPTIONS	Commande, gestion des prêts, facturations, relances. Gestion des formats PDFet EPSF. Option multithesaurus avec Ajaris-Gold	Aucune option proposée

IV. Analyse des fonctionnalités documentaires de plusieurs banques d'images en ligne

Nous avons procédé à une analyse de 8 banques d'images en ligne. Nous avons choisi ces banques on-line en fonction de leur taille (certaines sont des petites agences photographiques, d'autres de vastes portails), de leur logiciel, de leurs langages documentaires.

La méthode d' analyse de ces sites s'est structurée autour des questions suivantes :

- Qui : qui produit ce site et à qui est-il destiné ? Sur internet des ressources très spécialisés côtoient des ressources destinées à un public plus large et ceci vaut également pour l'image. Il est donc nécessaire de se demander si la source vise le grand public, les professionnels ou les scientifiques.

- Quoi : quelle est la nature du fonds, combien de documents possède t-il en ligne?

- Où : d'où proviennent les informations obtenues ? Qui est responsable du légendage, de l'indexation ?

- Quand : quelle est la date de création du fonds, mais aussi la date de création du site. Ces informations sont souvent indiquées dans la page et parfois dans les méta-données.

- Comment : Quelle est l'accessibilité du site ? Quelles sont les fonctionnalités de recherche, de téléchargement , … ?

- Pour quoi : quelles sont les fonctionnalités de commande ?

1.Grilles d'évaluation des banques d'images

SITE	IMAGE FORUM	AKG	CORBIS
adresse du site	http://www.imageforum.par.afp.com/	http://www.akg-images.com	http://pro.corbis.com
auteur	portail rassemblant l'AFP + d'autres agences (Roger Viollet, PPCM, Pictor, Rex…)	Archiv fur Kunst und Geschischte	plate forme de distribution des 3000 fonds d'archives rachetés par Corbis (news : Kipa, Sygma, Outline, Tempsport … et histoire /illustration : Bettman, National Gallery, Turnley, Ansel Adams, Musée du Louvre, de l'Ermitage,,,,)
date de création du fonds	1958 pour le fonds de l'AFP, 1999 pour ImageForum	1945, 2003 pour la mise en ligne	1990
nature du fonds	Actu, Archives, People, illustration, sport	beaux-arts, histoire, patrimoine, archéologie, civilisations, religions, musique, personnages historiques et mythologiques, photographies anciennes, cinéma.	histoire, beaux arts, actualités, sport, célébrités, portraits, illustration
accessibilité aux images	accès par loggin	accès par loggin	accès avec loggin pour agrandir et télécharger des images sans marquage, accéder à une notice, constituer un panier
contenu visible sans loggin	rien	rien	tout le reste
public visé	professionnels	professionnels	professionnels et grand public
gestion des droits d'auteurs	oui	oui	système complexe entre photos libres de droits, restrictions géographiques,…
aide en ligne	oui	oui	oui
version etrangère	anglaise, espagnole, allemand, portugais	langue pivot : allemande ; traductions en français et en anglais	allemand , anglais, espagnol, japonais, portugais (mais recherche en français ou en anglais uniquement)
analyse du fonds iconographique			
nombre de photos en ligne	2, 6 millions	450 000	3 millions dont 60 000 photos libres de droits
format de l'image en ligne	vignette, grand format basse résolution, grand format haute résolution	vignette, basse définition, moyenne définition, haute définition	basse, moyenne, haute, ultra
existence de watermark	non	oui	oui dès l'agrandissement ou le téléchargement
alimentation	500 photos par jour	une centaine par jour	?
analyse du traitement documentaire de l'image			
responsable légendage image	journalistes pour la partie légende	documentalistes et historiens d'art	variable selon fonds actu (1/3) et fonds illustration (2/3)
description de l'image	description très complète (date, crédit, photographe, source, ville, pays, référence, légende en anglais) lisible en cliquant sur le "slug" placé sous l'imagette	description complète avec année de l'œuvre, année de l'évenement, référence, crédit, et mots clé (avec lien hypertexte relancant automatiquement une interrogation sur les photos indexées avec le même mot clé)	informations sur n° référence, photographe, date de prise de vue, lieu.en cliquant sur l'image : informations sur les mots clé de la photo, les restrictions lié aux droits d'auteurs, la disponibilité de la haute definition
présentation des modes de recherches	Recherche simple dans un ou plusieurs champ. Recherche par descripteurs.Recherche avancée par champs spécifiques.Recherche avec opérateurs booléens, troncatures et caractères manquants.Historique de la recherche.	Recherche sur texte intégral. Recherche par descripteurs. Recherche par mots clé avec affichage des mots clés existants et possibilité de relancer une recherche sur l'un des mots clé. Recherche avec opérateurs booléens et troncatures.Historique de la recherche.	Recherche en texte intégral. Recherche par descripteurs. Affichage des mots clé de la photo et relance de recherche sur un ou plusieurs mots clé de la photo
évaluation du langage documentaire	Très peu d'analyse connotative. Thesaurus. Certains termes sont indexés en anglais donc parfois compliqué pour la recherche.	Analyse morphologique de l'image mais peu d'analyse connotative.Thésaurus en allemand avec gestion des traductions françaises et anglaise. Listes d'autorité.	Beaucoup d'analyse connotative. Thesaurus. Listes d'autorité.
modes de presentation des résultats	Mode mosaique ou mode liste .	Mode mosaique ou mode liste paramétrables par l'utilisateur. Une liste de mots clé associés à sa requête s'affiche dans la partie gauche de l'écran.	Mode mosaique avec affichage de la notice documentaire et de la disponibilité de l'image. Possibilité d'afficher la photo dans son contexte de reportage
mode de commande en ligne	Téléchargement de l'image en haute définition	Téléchargement de l'image en haute définition	possibilité de télécharger l'image maquette sans loggin ou de commander l'image haute défintion avec envoi par e-mail du fichier compressé
logiciel utilisé	Imageforum	Orphea de la société Algoba	Image Seeker + logiciel maison

SITE	CIRIC	LUMIERE DU MONDE	MAGNUM PHOTO
adresse du site	www.photociric.com	http://www.lumieredumonde.com	http://www.magnumphotos.com
auteur	agence Ciric	agence photographique Lumière du monde	Agence Magnum
date de création du fonds	1945, 1992 pour la mise en ligne, 2004 pour la nouvelle mise en ligne	2002	1947, 2002 pour la mise en ligne
nature du fonds	agence d'illustrations et de reportages consacrée à l'information religieuse, sociale et humanitaire	images d'illustrations et d'actualité sur le thème du religieux	agence de reportages
accessibilité aux images	accès par loggin	accès avec loggin pour la recherche avancée et le téléchargement d'images	accès par loggin pour la recherche uniquement
contenu visible sans loggin	rien	50 reportages d'actualités, 30 sélections	énormément de reportages, de webfolios de photographes, de sélections de livres, d'expositions
public visé	professionnels	professionnels	professionnels et grand public
gestion des droits d'auteurs	oui	oui	oui
aide en ligne	oui	oui	oui
version etrangère	non	non	anglaise
analyse du fonds iconographique			
nombre de photos en ligne	45 000	20 000	450 000
format de l'image en ligne	vignette, moyenne et haute résolution	vignette, moyenne et haute résolution	vignette,basse et haute définition
existence de watermark	non	oui	oui
alimentation	100 par semaine	très variable- actuellement aucune	entre 500 et 800 par jour
analyse du traitement documentaire de l'image			
responsable légendage image	photographes et documentalistes	photographes	documentalistes
description de l'image	informations sur date de création, titre, légende, réf., stock, signature, mots-clés (avec lien hypertexte relancant automatiquement une interrogation sur les photos indexées avec le même mot clé)	notice documentaire avec référence de la photo, titre, légende, date, pays, auteur, mots clé	notice documentaire avec légende (en anglais), réf. de la photo, nom du photographe, sélection dont elle est tirée. possibilité d'afficher ou de masquer les mots clés (en français) (avec lien hypertexte relancant automatiquement une interrogation sur les photos indexées avec le même mot clé)
présentation des modes de recherches	Recherche en texte intégral. Recherche par descripteurs.Recherche par mots clé. Recherche avec opérateurs booléens, troncatures. Historique de la recherche.	recherche par mots clé ou par thème avec accès au plan de classement. Affichage des mots clé. Opérateurs booléens Historique de la recherche.	Recherche intégrale. Recherche par descripteurs. Recherche par mots clé. Recherche sur les sélections de magnum depuis 1997.Historique de la recherche
évaluation du langage documentaire	Beaucoup de surindexation.Peu d'analyse morphologique ni connotative. Listes d'autorité.Pas de thesaurus	plan de classement, listes d'autorité.Thesaurus	Listes d'autorité.Thesaurus bilingue .
modes de presentation des résultats	Mode mosaique ou mode liste paramétrables par l'utilisateur. Possibilité d'afficher la photo dans son contexte de reportage. Une liste de mots clé associés à sa requête s'affiche dans la partie gauche de l'écran.	Mosaïque avec de 2 à 64 images par page.	mosaique. Affichage par tri (les plus anciennes, les plus récentes,les meilleures, par photographe). Affichage des mots clé de la photo possible avec renvois vers d'autres recherches.
mode de commande en ligne	Téléchargement de l'image en ligne sous ses différentes définitions	envoi par mail des photos sélectionnées (soit vignettes soit hautes définitions)	envoi de photo en haute définition par email
logiciel utilisé	orphea	logiciel maison	Teldin puis Cortext (orange logic)

Nom du site	RUE DES ARCHIVES	ROGER VIOLLET
adresse du site	http://www.rue-des-archives.com/	http://www.roger-viollet.fr/principale.asp
auteur	agence rue des Archives	agence Roger Viollet
date de création du fonds	1992 pour l'agence, 2000 pour la mise en ligne	1998
nature du fonds	très diversifié (vie politique,sociale, culturelle française des années 30 à 80+ fonds américain)	archives historiques sur l'Europe et différents pays du monde + Art
accessibilité aux images	accès par loggin pour la recherche uniquement	accès par loggin pour recherche
contenu visible sans loggin	sélections thématiques (Rétro et actualités) et vitrine de l'agence à partir de 7 thèmes (jazz, insolite, cinéma, histoire …)	toutes les sélections
public visé	profesionnel	professionnels
gestion des droits d'auteurs	?	oui
aide en ligne	oui	oui
version etrangère	anglaise (traduction anglaise des mots clé s'affichant a la suite des mots clé français, pas du tout pratique)	non
analyse du fonds iconographique		
nombre de photos en ligne	120 000	70 000
format de l'image en ligne	vignette , mosaïque, zoom en basse défintion, document original (haute définition)	vignette, zoom en basse définition, haute définition
existence de watermark	oui	non
alimentation	200 par semaine	100 semaine
analyse du traitement documentaire de l'image		
responsable légendage et indexation	documentalistes	documentalistes
description de l'image	notice comprenant titre, légende, copyright, références, sources, mots clé (avec lien hypertexte relancant automatiquement une interrogation sur les photos indexées avec le même mot clé)	notice très succinte (uniquement légende, date et copyright)
présentation des modes de recherches	Recherche sur texte intégral Recherche par descripteurs Recherche par mots clé avec affichage des mots clés associés. Recherche avec opérateurs booléens et troncatures. Historique de la recherche.	Recherche simple (avec années, couleurs, mots clés). Recherche par descripteurs.
évaluation du langage documentaire	Listes de mots clé associés. Thesaurus	Listes mots clé. Listes d'autorité. Pas de thesaurus
modes de presentation des résultats	Mosaïque avec affichage de la notice documentaire et de la disponibilité de l'image. Une liste de mots clé s'affiche dans la partie gauche de l'écran.	Mosaïque
mode de commande en ligne	possibilité de télécharger l'image (zoom ou haute définition) sur loggin	possibilité de télécharger l'image (zoom ou haute définition)
logiciel utilisé	orphea produit par la société Algoba	fuldesk de la société Fulcrum

2. Résultats et surprises

Ce qui ressort en premier lieu de cette analyse c'est une certaine homogénéité des fonctionnalités documentaires malgré de grandes disparités dans les structures des agences. Cette homogénéité semble liée au développement des logiciels spécialisés dans l'image ; des agences de taille très différentes ont acquis le même logiciel et ont une interface web très proche (par exemple Akg avec 450 000 images numérisées et Ciric avec 10 fois moins d'images).

Les similitudes :

- l'affichage
- la partie éditoriale (beaucoup de logiciels ont intégré une fonction éditoriale : sélection, reportages, portfolios...)
- une grande facilité pour le catalogage (toutes les agence visitées ont importé le format IPTC comme format standard)
- les modes de recherche (recherche simple, avancée)
- l'aide en ligne, de par la non-relation directe avec les utilisateurs

Les disparités :

- l'indexation (très variable selon les agences)
- la mise en ligne de leurs langages documentaires : certaines agences fonctionnent en interne avec un langage contrôlé, un thesaurus, ou un langage classificatoire, mais ces langages ne sont pas proposés en ligne (les sites Gamma, Sipa ou Magnum travaillent avec des listes de mots- clés associés, des thésaurus qui ne sont pas visibles sur leur interface) [annexes 6 et 7]. D'autres agences proposent leurs listes de mots- clés et leurs mots-clés associés, [annexe 8], d'autres affichent leur thésaurus [annexe 9].

Quoiqu'il en soit on peut souligner qu'avec l'émergence des réseaux informatiques et avec la possibilité pour l'utilisateur final d'accéder aux banques de données, les langages documentaires sont devenus plus conviviaux, les vocabulaires contrôlés plus proches du langage naturel des utilisateurs.

V. Tendances

1. L'émergence de portails

Pour diminuer le temps de recherche on voit émerger de plus en plus des portails d'agences photographiques qui se regroupent. [annexe n°10]

Les iconographes ont en effet avoir de moins en moins de temps à consacrer à la recherche d'images. Les directeurs photos de la presse sont connectés à une cinquantaine de sites en moyenne par jour avec une très nette tendance à se focaliser sur les 5 agences les plus importantes : AP, AFP, Reuters, Getty et Corbis. Pour gagner du temps ils effectuent leurs recherches iconographiques sur des portails qui regroupent plusieurs fonds et donc plusieurs langages documentaires, entraînant une hétérogénéité des langages. Or on constate que la méthode de regroupement des vocabulaires contrôlés est d'autant plus complexe que les outils linguistiques sont déjà construits et opérationnels. « Peut-on fusionner ou même faire cohabiter des outils linguistiques spécifiques à un groupe donné quand on sait à quel point le langage peut être révélateur d'une identité et d'une culture ? » (Compatibilité des langages d'indexation. Mariage, fusion ou cohabitation ?, p.3) [19] Pourtant ce problème de compatibilité entre langages est crucial aujourd'hui puisque la crise économique tend à regrouper les moyens, les techniques et les langages. D'autant plus que le bilinguisme des thésaurus est devenu une nécessité en raison de l'internationalisation des banques d'images et des clients.

Certaines sociétés y travaillent aujourd'hui, comme Hachette Filipacchi Medias. Le groupe prépare l'ouverture de son portail avec le logiciel Everteam (groupe Ever), qui regroupera toutes les agences du groupe, dont Gamma, Keystone, Rapho, Hoa Qui, Explorer, Jacana, Stills, Top, Katz ect. Un très gros travail de compatibilité de langages documentaires est actuellement mené par des documentalistes du groupe, qui avaient déjà élaboré des outils d'indexation communs pour les agences Rapho, Explorer, Jacana et Hoa-Qui lorsque celles-ci avaient emménagé dans les mêmes bureaux en 2000. La récupération d'autres outils documentaires (comme ceux de Gamma par exemple, très opérationnels) va nécessiter une mutualisation de thésaurus, de listes d'autorité, de tables de contrôle.

2. L'indexation automatique

Bien que développées tardivement par rapport à la croissance exponentielle de l'audiovisuel dans notre société, les recherches en indexation automatique se multiplient depuis quelques années.
L'indexation automatique est effectuée exclusivement par des moyens informatiques et n'implique à aucun moment l'utilisation du langage. Elle consiste à extraire un certain nombre de paramètres d'analyse d'images qui serviront à la caractériser. Elle s'effectue par l'extraction et la représentation numérique d'attributs significatifs de l'image sous forme d'index appelés « signatures » (forme, couleur, texture de l'image, reconnaissance d'objets d'un type donné comme des visages, des empreintes digitales, …) L'utilisateur fournit ou choisit une image-requête se rapprochant de sa recherche et le système retrouve les images dont les résultats d'analyse s'apparentent à ceux de l'image–requête. La recherche peut s'effectuer en utilisant une signature ou une combinaison de signatures. Ces méthodes ont été mises en place pour des sources d'images homogènes, et pour des domaines précis (Etude comparative des fonctionnalités de moteurs de recherche d'images sur internet, p. 98) [14]

Dans plusieurs applications ces outils d 'indexation automatique viennent en complément des outils sémantiques existants. C'est le cas dans les applications suivantes :

- Le système d'indexation automatique basée sur le contenu de l'image, ou CBIR (content based-image retrival) est utilisé par le programme IMEDIA développé par l 'INRIA ((Institut National de Recherche en Informatique et Automatique). Le logiciel d'analyse de contenu et de recherche d'images fixes appelé Surfimage, permet de rechercher un certain type d'images au sein d'une base de données grâce à un affinage progressif et interactif de la requête[12]. Ses domaines d' applications sont les suivants :

  - Internet (images et vidéos),
  - Audiovisuel (personnage dans les JT, documentaires, sport, …)
  - Médecine (recherche à but diagnostic ou pédagogique)
  - Art et Design (archives archéologiques, peintures, tissus, …)
  - Authentification (visages, empreintes digitales, logos)
  - Sécurité (surveillance vidéo, objets d'arts volés …)
  - Education (recherche encyclopédique)

- La société LTU Technologies, éditeur de logiciels d'analyse et de reconnaissance des images, a déposé en 1999 un brevet international s'appliquant aux photographies, dessins, illustrations ou tout autre document visuel numérisé, permettant d'indexer, de reconnaître et de comparer les images à partir de leurs composantes visuelles. Elle a réalisé un prototype

de recherche par l'image, en partenariat avec Corbis, qui est opérationnel sur plus de 65 000 images libres de droits. A la suite d'une requête textuelle, les utilisateurs peuvent sélectionner la fonction "Find similar images" pour immédiatement trouver des images visuellement similaires. Image-Seeker est un logiciel permettant la recherche et la navigation par similarité visuelle dans des collections d'images et fonds iconographiques. Chaque image est ainsi indexée à partir de ses formes, couleurs, textures, configurations spatiales des éléments etc. Cela crée un vecteur (appelé ADN par LTU Technologies) auquel est ajoutée la prise en compte des métadonnées textuelles provenant de l'annotation manuelle opérée par Corbis et ses documentalistes. La similarité entre 2 images est donc une combinaison des éléments visuels et des concepts associés. [annexe 11]

Mais les capacités réelles de cette technique en matière de reconnaissance formelle sont encore limitées, et ses applications semblent inopérantes en photo-journalisme et difficiles à mettre en place également pour l'illustration, les scènes à analyser étant complexes et subjectives.

Ces recherches sur l'indexation automatique des images connaissent surtout un net développement dans le domaine de l'audiovisuel. Plusieurs programmes ont été développés, tant en France (à l'Institut National de l'Audiovisuel notamment) qu'au niveau européen avec les projets EUROMEDIA, DIVAN, AVIR, COBWEB, ou STRECH.

Conclusion

Après plusieurs années au cours desquelles on a vu se multiplier les banques d'images, la tendance est plutôt à une régulation du marché. Certaines agences se sont regroupées, d'autres se sont spécialisées, des petites ont disparu. Les sociétés semblent avoir désormais comme objectif principalement économique, une meilleure qualité des images et un meilleur service au client. En effet, si disposer immédiatement de l'image recherchée est devenu un rêve quasi réalisable, encore faut-il pouvoir trouver cette image idéale dans l'immense choix disponible sur Internet.

Pour faire face au défi technologique de la maîtrise d'accès aux images, nous avons observé qu'une construction astucieuse de langages documentaires restait un garant de succès. Le choix de méthodes efficaces d'indexation et de catalogage (avec le format IPTC), associées à l'affichage des images dans leur contexte, comme le proposent aujourd'hui tous les logiciels du marché, permet une recherche iconographique optimale et concerne aussi bien les petites agences que les grosses structures ou les portails. Mais ce choix nécessite du temps, de l'argent et une vraie connaissance de la documentation, qui n'est absolument pas de mise dans certaines banques d'images qui tablent avant tout sur le stockage et sur les progrès de l'indexation automatique. L'indexation d'images par le contenu a l'avantage de permettre le choix d'une image en la comparant à d'autres images similaires et peut s'appliquer aisément le cas de petites collections regroupant des images simples, mais elle reste difficilement envisageable dans le monde de la presse et de l'illustration. et ne peut en tout cas l'être qu'associée à de solides outils sémantiques existants.

Soulignons enfin la difficulté à faire émerger le traitement intellectuel réalisé en interne, difficulté essentiellement due au manque de lisibilité des sites des agences, qui sont souvent conçus en fonction de la communication plutôt que de la documentation.

Bibliographie

Bibliographie analytique

Cette bibliographie arrêtée le 26.09.2004 contient 38 références classées autour de grand thèmes. A l'intérieur de chaque thème les références bibliographiques sont classées par ordre chronologique inverse.

Liste des thèmes :
Ouvrages de référence, p. 68
Généralités sur l'image fixe, p. 68
Traitement de l'image fixe et langages documentaires, p. 69
La démarche qualité, p. 72
Actualités du marché de l'image en ligne, p.73

Les numéros devant les références correspondent aux numéros entre crochets accompagnant les citations dans le corps du mémoire.

Ouvrages de référence

1 *Vocabulaire de la documentation*, sous la direction d' A. Boulogne. Paris : ADBS éditions, 2004, 335 p.

2 Maniez, J. *Actualité des langages documentaires : fondements théoriques de la recherche d'information*. Paris : ADBS éditions, 2002, 396 p.

Dans cet ouvrage, l'auteur analyse le rôle et l'importance que conservent aujourd'hui les langages d'indexation et de recherche, à l'ère d'Internet. Il y détaille les composantes fondamentales des langages documentaires, les langages classificatoires hiérarchiques , les langages syntagmatiques et les langages à base de descripteurs ainsi que les problèmes de l'indexation.

3 Cacaly S., Le Coadic Y.F, Melot M., Pommart P.D, Sutter E. *Dictionnaire encyclopédique de l'information et de la Documentation*. Paris : Editions Nathan, 1997, 664 p.

Définition des concepts fondamentaux et description des outils techniques, technologies et méthodes. Articles sur l'indexation p. 284, l'image p. 277, le thésaurus p. 577.

Généralités sur l'image fixe

4 *Le répertoire Iconos (Photohèques et photographes*). Paris : La Documentation française (Coll. Photodoc) , 2004, 575 p.

5 Kattnig, C. *Gestion et diffusion d'un fonds d'image*. Paris : Editions Nathan (Coll. Lettres), 2002, 127 p.

Cet ouvrage méthodologique décrit les différentes étapes de la mise en place d'une banque d'images en s'attachant aux aspects techniques et juridiques et au traitement documentaire.

6 *Nouvelle histoire de la photographie*, sous la direction de M. Frizot. Paris : Larousse, 2001, 776 p.

Ouvrage général très bien documenté retraçant la chronologie de la photographie et ses diverses thématiques.

7 Klinkenberg, J.M. *L'univers des signes*. *Sciences Humaines*, mai 1998, n° 83, p.20-23

Article de fond sur la sémiotique, comportant un passage consacré à l'image.

8 Joly,M. *L'invisible dans l'image*. *Sciences Humaines*, 1998, n° 83, p.26-29.

Analyse sémiologique et lecture de l'image fixe

9 Gervereau, L. *Voir comprendre, analyser les images*. Paris : La Découverte,1997, 192 p.

Guide d'analyse des images proposant des méthodes et une grille d'analyse pour tous les types d'iconographie, avec de nombreux exemples.

10 Joly, M. *L'image et les signes : approches sémiologiques de l'image fixe* . Paris : Editions Nathan (Coll. Images), 1994, 191 p.

Etude de l'analyse et de la signification de l'image fixe d'un point de vue sémiologique.

11 Barthes, R. *Rhétorique de l'image*. *Communications*, 1964, n°4, p. 40-51

Article dans lequel Barthes pose les premiers jalons d'une sémiologie de l'image à partir de l'analyse d'une image publicitaire de Panzani : l'image est composée de différents types de signes : linguistiques, iconiques, plastiques, dont résulte une signification globale et implicite.

Traitement de l'image fixe et langages documentaires

12 Indexation et recherche par le contenu visuel dans les documents multimédia. [consulté le 29/09/2004] http://savoirscdi.cndp.fr/rencontrelyon/boujemaa/boujemaa.pdf

Nozha Boujemaa, Chef du projet IMEDIA à l'Institut National de Recherche en Informatique et en Automatique, présente ses travaux sur l'indexation par le contenu visuel.

13 International Press Telecommunications Council [consulté le 29/09/2004] http://www.iptc.org/pages/about_main.php

Pour tout savoir sur l'histoire et les applications du format IPTC.

14 Boudry, C et Agostini, B. *Etude comparative des fonctionnalités de moteurs de recherche d'images sur internet*. *Documentaliste Sciences de l'information*, 2004, vol. 41, n°2, p.96 – 105

L'objectif de cet article est donc de comparer les fonctionnalités de ces outils de recherche, en distinguant ceux qui dérivent directement d'outils de recherche de pages web de ceux qui ont été spécialement créés et développés pour la recherche d'images sur inernet.

15 En quoi les nouvelles technologies renouvellent-elles le traitement intellectuel de l'image ? [consulté le 16/09/2004] http://savoirscdi.cndp.fr/rencontrelyon/kattnig/kattnig.ppt

L'auteur analyse les apports des nouvelles technologies pour la gestion d'un fonds d'images plus précisément en matière de reproduction et de diffusion, et détaille la démarche d'un documentaliste dans le contexte numérique.

16 André , S. *Journée d'étude ADBS, Images numériques : quels formats, pour quels usages* ? *Documentaliste - Sciences de l'information*, 2003, vol.40, n°6, p.382-386

Pour le point sur les bases techniques de l'image numérique, retenu ici pour le chapitre sur le numérique et les fonds photographiques.

17 *La gestion des archives photographiques,* sous la direction de N.Charbonneau et M.Robert. Sainte Foy : Presse de l'Université du Québec (Gestion de l'Information), 2001, 326 p.

Les auteurs traitent de manière complète la gestion des archives photographiques sous des angles à la fois théoriques et pratiques (histoire et techniques de la photographie, acquisition, préservation et diffusion …) Nous retiendrons surtout la partie consacrée à la classification, à l'analyse de contenu, à la description et à l'indexation.

18 Clerté, J. *Etude comparative du traitement documentaire des banques d'images sur internet*. Mémoire INTD n°30-15, 2000, 71 p.

Evaluation des différentes méthodes d'analyse et de présentation de l'image sur Internet, à travers les problématiques liées au traitement de l'image et comparaison de plusieurs banques d'images en ligne.

17 Moszer, S. *Les spécificités du traitement documentaire de l'image : Harmonisation d'indexation à l'agence BDDP Interactive*. Mémoire INTD n°28- 47,1998, 82 p.

Ce mémoire a pour but de développer les spécificités du traitement documentaire intellectuel de l'image fixe et animée. La première partie fait état de quelques réflexions théoriques sur le document image, les moyens à mettre en oeuvre pour résoudre les problèmes que posent l'analyse de l'image, ainsi que les avancées des recherches menées sur les nouveaux procédés d'analyse de l'image.

19 Degez, D. *Compatibilité des langages d'indexation. Mariage, cohabitation, fusion ? Quelques exemples concrets*. *Documentaliste Sciences de l'information*, 1998, vol.35, n°1, p.3-14

Etude des différents problèmes à résoudre pour la mise en compatibilité des langages documentaires.

20 *Norme NF Z 44-077 : Documentation. Catalogage de l'image fixe . Rédaction de la description bibliographique***.** Paris : AFNOR, 1997, 179 p.

Le présent document expose les règles relatives à la description bibliographique des images fixes, précise un ordre pour les éléments de la description et codifie la ponctuation. La partie retenue ici concerne les images fixes et la description bibliographique des images fixes dans des catalogues multimédia.

21 Auclair, A. *Elaboration d'un outil linguistique à l'usage d'une banque de données d'images : le cas Hachette*. Mémoire INTD n°27-02, 1997, 159 p.

L'étude porte sur la création d'un outil linguistique (une liste hiérarchique) dans le cadre d'un projet de banque de données images dirigé par Hachette-Livres pour gérer sa photothèque.

22 Guérin, M.A. *Etude sur le langage documentaire d'une banque de données iconographiques*. Mémoire INTD n°27-35, 1997, 109 p.

Mémoire sur les problèmes de l'analyse documentaire des images fixes au sein d'une agence photographique et plus particulièrement les problèmes de connotation.

23 Henry, P. *Gestion de bases de données : l'indexation automatique d'images. Technologies internationales*, 1997, n° 39, p.7-11

Article qui évoque les avancées en matière d'indexation automatique des images.

24 Guilbaud, E. *Comment indexer l'image fixe ? Archimag*, 1995, n°86, p.33-36

L'auteur analyse des différences entre l'indexation d'un fonds photos classique et celle d'un fonds informatisé.

25 Collard C., Giannattasi I., Melot M*. Les images dans les bibliothèques*. Paris: Edition du Cercle de la librairie, 1995, 390 p.

On retiendra la partie sur les étapes du traitement documentaire des images fixes en bibliothèques.

26 Lubkov, M*. Les logiciels de photothèque multimédia*. *Archimag*, 1995, n°82, p.54-56

Point sur les grandes familles de logiciels dédiés au traitement documentaire des images numérisées.

27 Dauzat, M. *Le thésaurus de l'image : étude des langages documentaires pour l'audiovisuel*. Paris : ADBS éditions (Coll. Recherches et documents), 1994, 94 p.

Etude comparative de 14 thesaurus images dans les domaines de la culture, de la presse et de quelques autres secteurs techniques.

28 *Le document Image*. *Archimag,*1992, n° 5, 40 p.

Etude sur les spécificités du traitement documentaire de l'image fixe

29 F. Rabitti P, Savino. *Automatic image indexation to support content based retrieval*. *Information Processing and management*, 1992, vol.28 n° 5, p. 547-565

Les fondements de la recherche en indexation par le contenu.

30 HUDRISIER, H. *L'Iconothèque : documentation audiovisuelle et banques d'images*. Paris : La Documentation française, 1983, 269 p.

Réflexion théorique sur la communication et la pratique documentaire audiovisuelle et leurs applications, ici pour la partie photohèques.

31 Bléry, G. *La mémoire photographique : étude de la classification des images et de leur contenu à l'aide de l'informatique*. Bulletin *Interphotothèque*, 1981, n° 41, p. 9 –33.

L'auteur aborde la question de la subjectivité de l'image et démontre à l'aide d' exemples les problèmes liés à la polysémie de l'image. Elle propose une méthode d'indexation à ;partir de mots abstraits.

La démarche qualité

32 Sutter, E. *Documentation, information, connaissances : la gestion de la qualité*. Paris : ADBS éditions, 2002, 327 p.

Ouvrage de référence sur la qualtié dans le domaine de la documentation.

33 Duflos*, A. Les critères d'évaluation des banques de données : la démarche qualité chez les professionnels de l'information électronique*. Paris : ADBS éditions, 1995, 146 p.

Ouvrage fondé sur une enquête auprès de producteurs de banques de données.

Actualités du marché de l'image en ligne

34 Bordet, M. *Bill Gates, deuxième banquier d'images mondial*. *Challenges*, [En ligne], 2001, n° 166 [consulté le 16/09/2004] http://lemagchallenges.nouvelobs.com/coulisses/art2.html

35 Guerrin, M. *Hachette affirme ses ambitions dans la photo de presse*. *Le Monde*, 2000, 1er septembre.

36 Guerrin, M. *Photoreporters, les illusions perdues*. *Le Monde*, 2000, 6 septembre.

37 Guerrin *,* M. *La guerre des images et le photojournalisme. Le Monde Interactif* [En ligne],1999 (10 juin), [consulté le 25/09/2004] http://www.infoguerre.com/article.php?sid=63

38 Amar, P.J. *Le photojournalisme.* Paris :Editions Nathan, 2000, 128 p.

Annexes

Annexe 1 : Plan de classement proposé par l'agence Lumière du monde

religion >> - **animisme**
- **bouddhisme**
- **christianisme**
- **hindouisme**
- **islam**
- **judaïsme**

religion >> **animisme** >> - **vie spirituelle**

religion >> **bouddhisme** >> - **art**
- **lieux**
- objets
- **personnages**
- religieux
- **symbole & représentation**
- **vie quotidienne**
- **vie spirituelle**

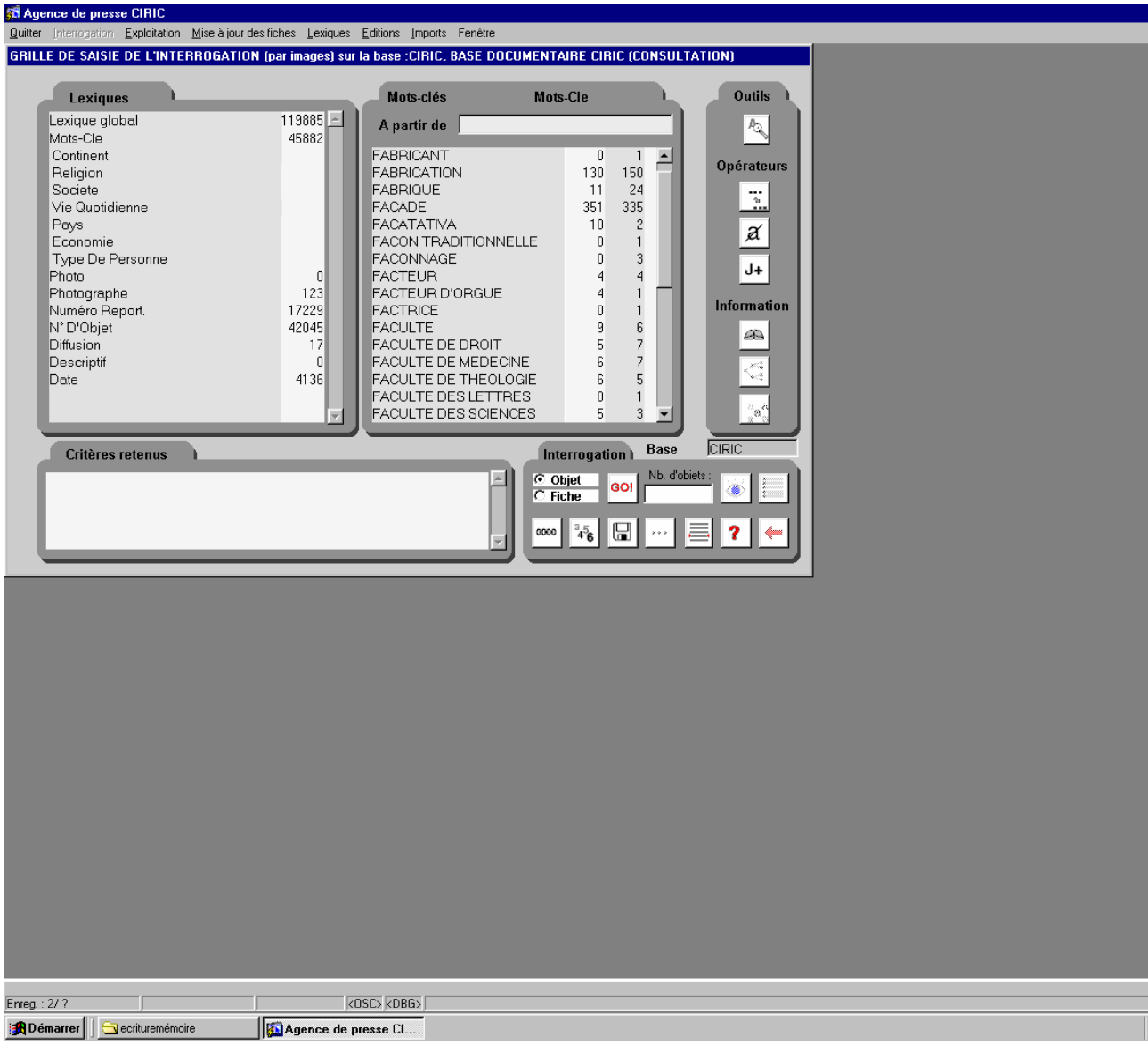
religion >> **christianisme** >> - **catholique**
- **courants & communautés**
- **orthodoxe**
- **protestant**

religion >> **hindouisme** >> - divers
- **lieux**
- **personnages**
- **vie spirituelle**

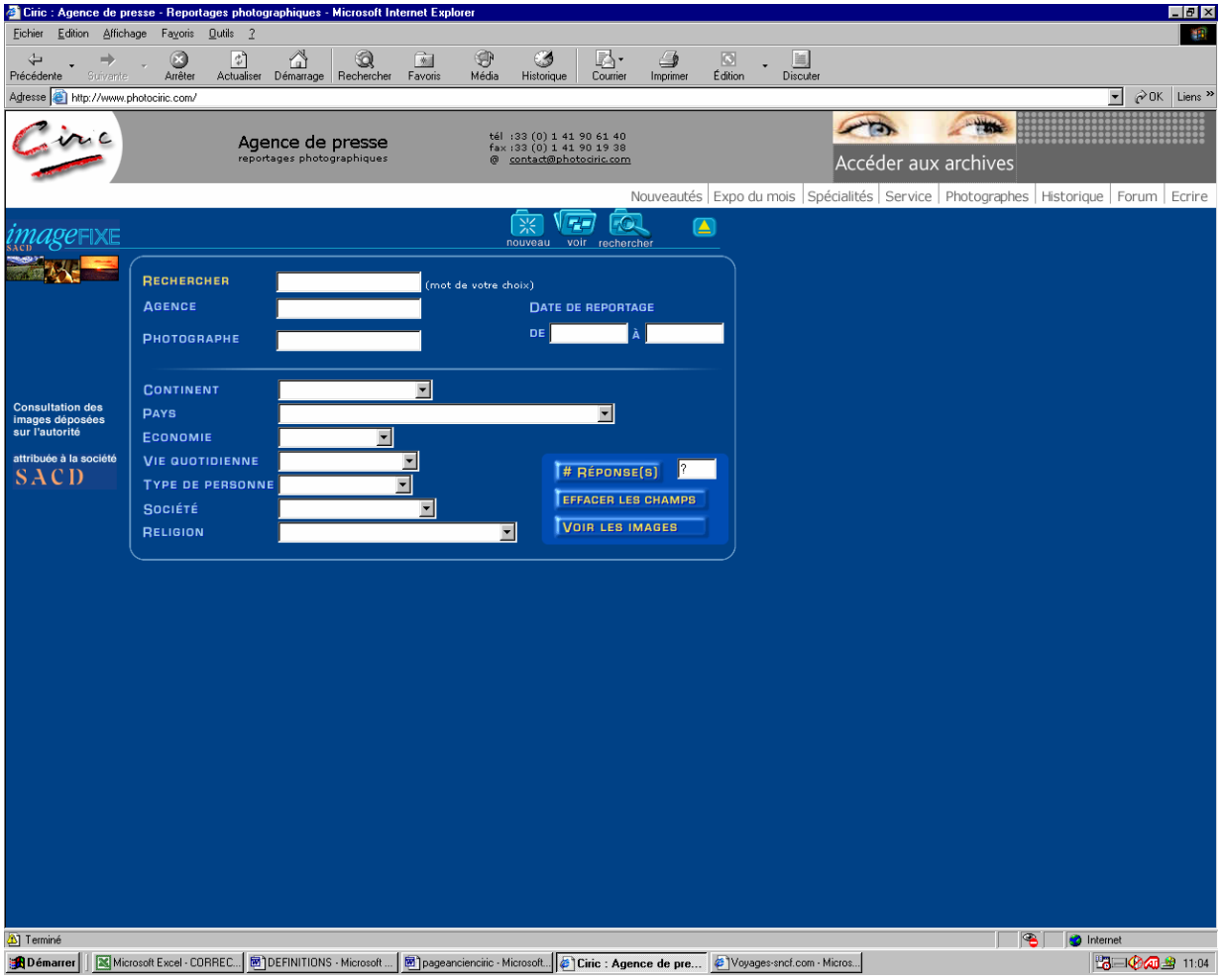
religion >> **islam** >> - **art**
- **fêtes religieuses**
- **lieux**
- religieux
- symbole & représentation
- textes sacrés
- **vie quotidienne**
- **vie spirituelle**

religion >> **judaïsme** >> - **courants & communautés**
- **fêtes religieuses**
- **lieux**
- objets
- religieux
- symbole & représentation
- textes sacrés
- vie quotidienne
- **vie spirituelle**

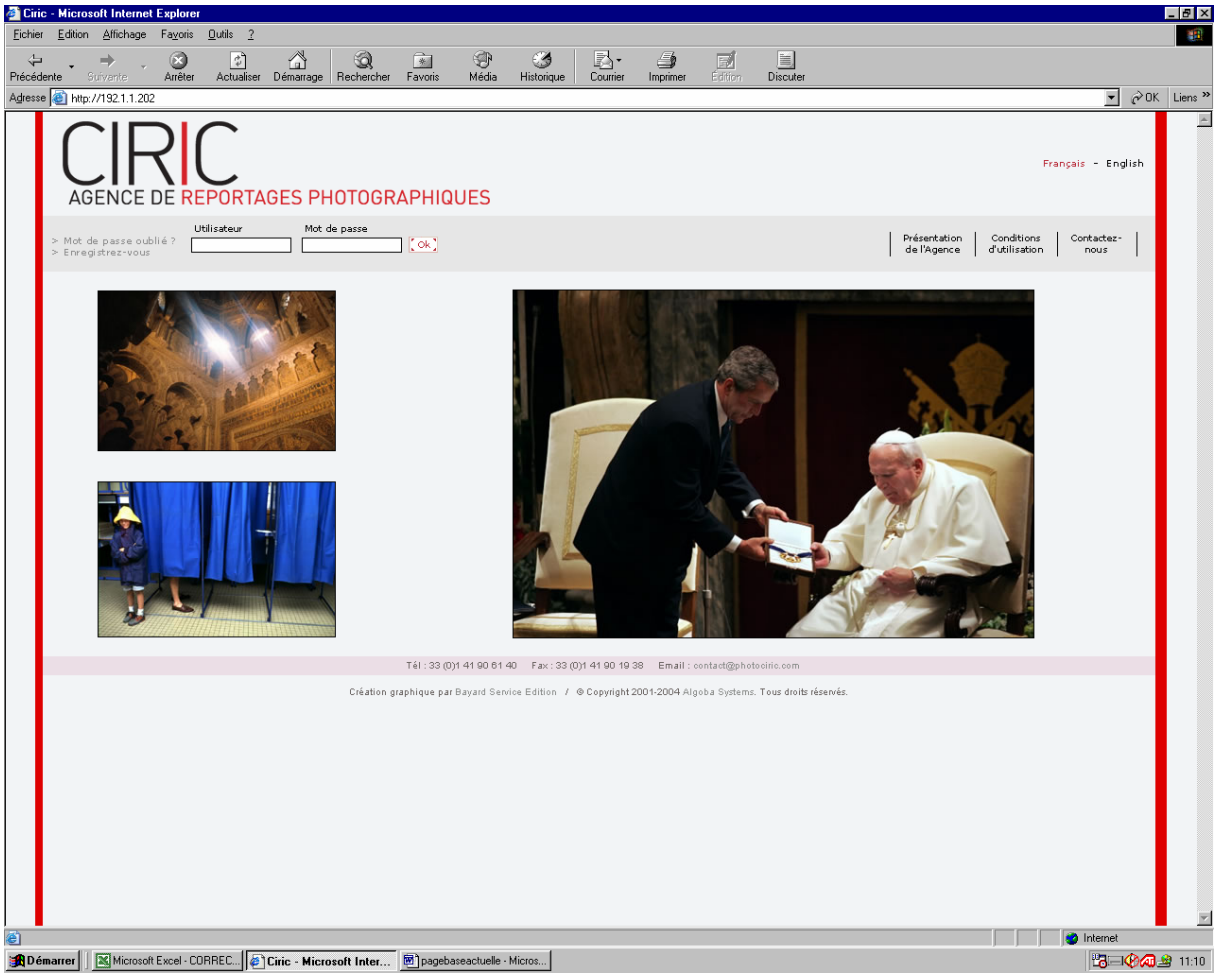
Annexe 2 : Interface de recherche sur l'ancienne base de Ciric (DFI)



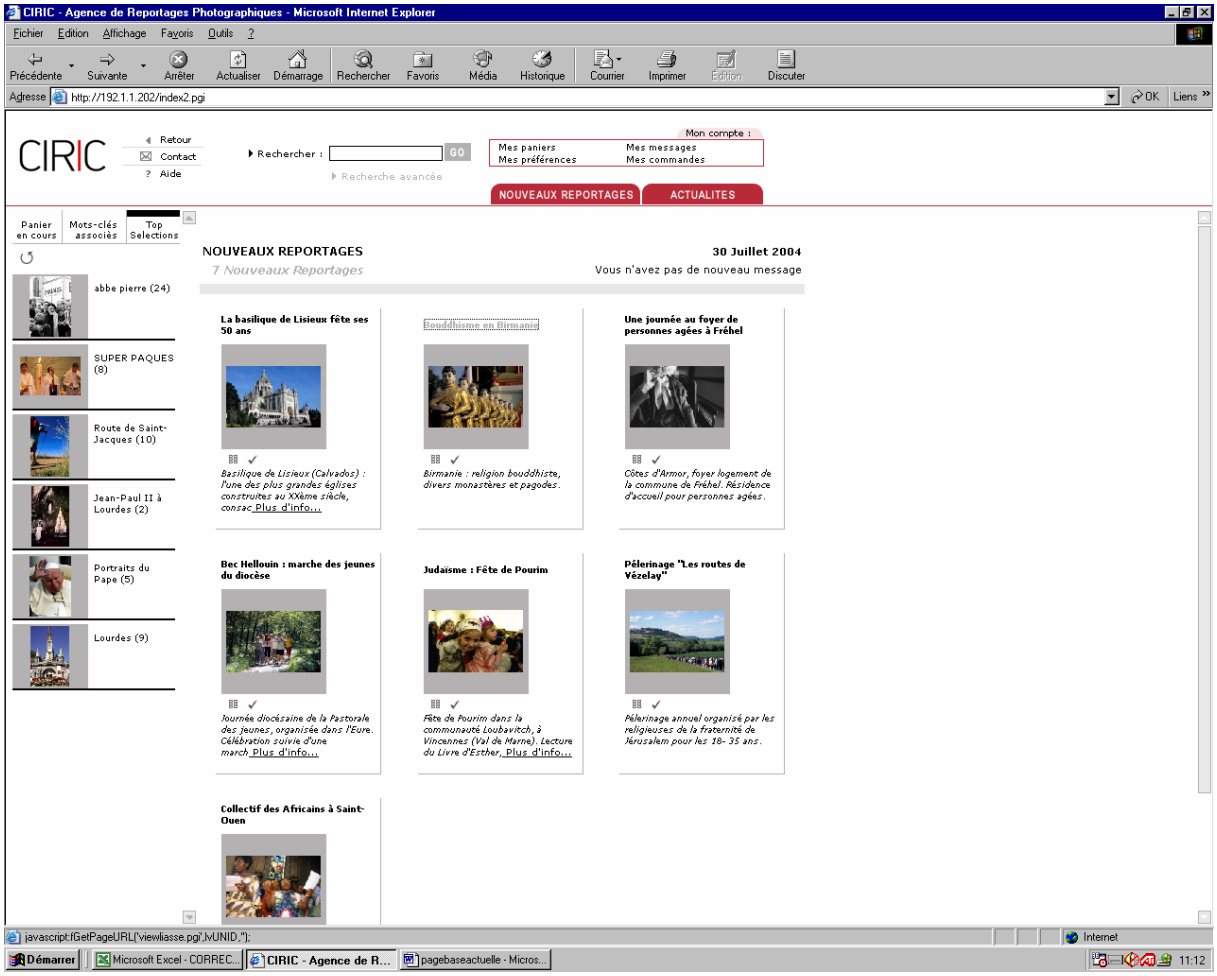
Annexe 3 : Interface de recherche sur l'ancienne interface web de Ciric (SACD)



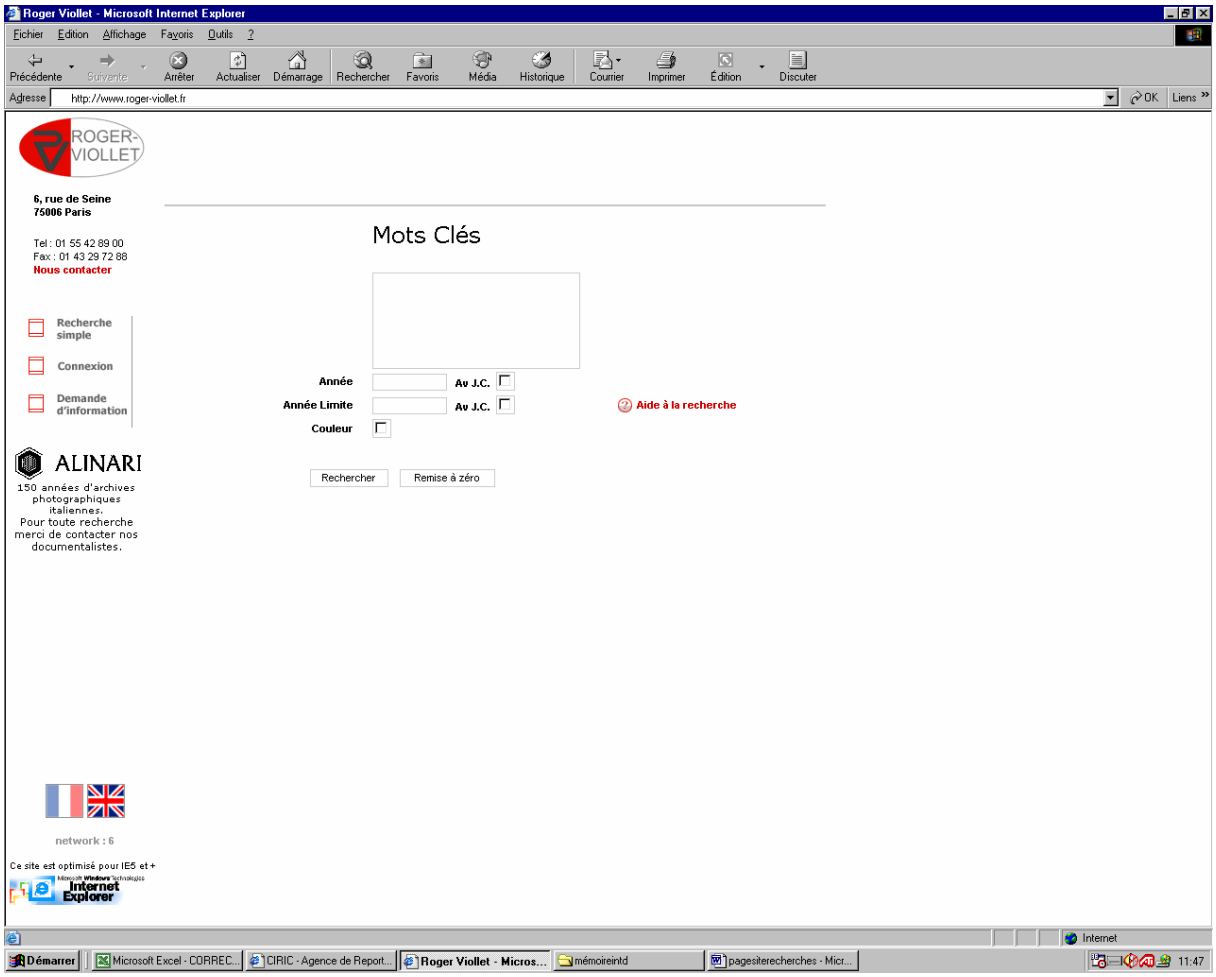
Annexe 4 : Interface web actuelle de CIRIC (page d'accueil)


Ciric - Microsoft Internet Explorer
Fichier Edition Affichage Favoris Outils ?
Précédente Suivante Arrêter Actualiser Démarrage Rechercher Favoris Média Historique Courrier Imprimer Edition Discuter
Adresse http://192.1.1.202
OK Liens
CIRIC
AGENCE DE REPORTAGES PHOTOGRAPHIQUES
Français - English
> Mot de passe oublié ?
> Enregistrez-vous
Utilisateur
Mot de passe
Ok
Présentation de l'Agence
Conditions d'utilisation
Contactez-nous
Tél : 33 (0)1 41 90 61 40 Fax : 33 (0)1 41 90 19 38 Email : contact@photociric.com
Création graphique par Bayard Service Edition / © Copyright 2001-2004 Algoba Systems. Tous droits réservés.
Internet
Démarrer
Microsoft Excel - CORREC...
Ciric - Microsoft Inter...
pagebaseactuelle - Micros...
11:10

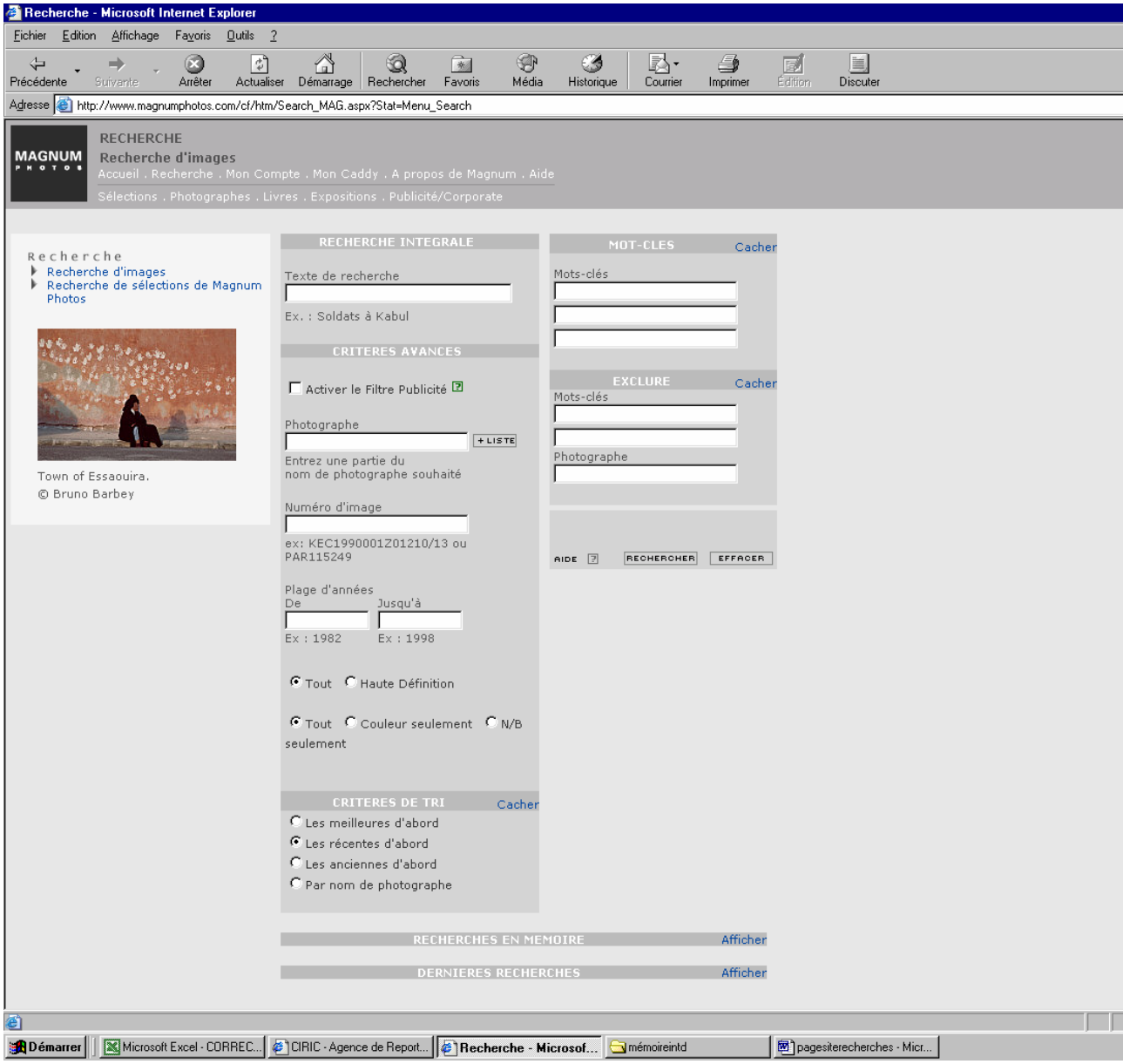
Annexe 5 : Interface web actuelle de CIRIC (niveau 1)



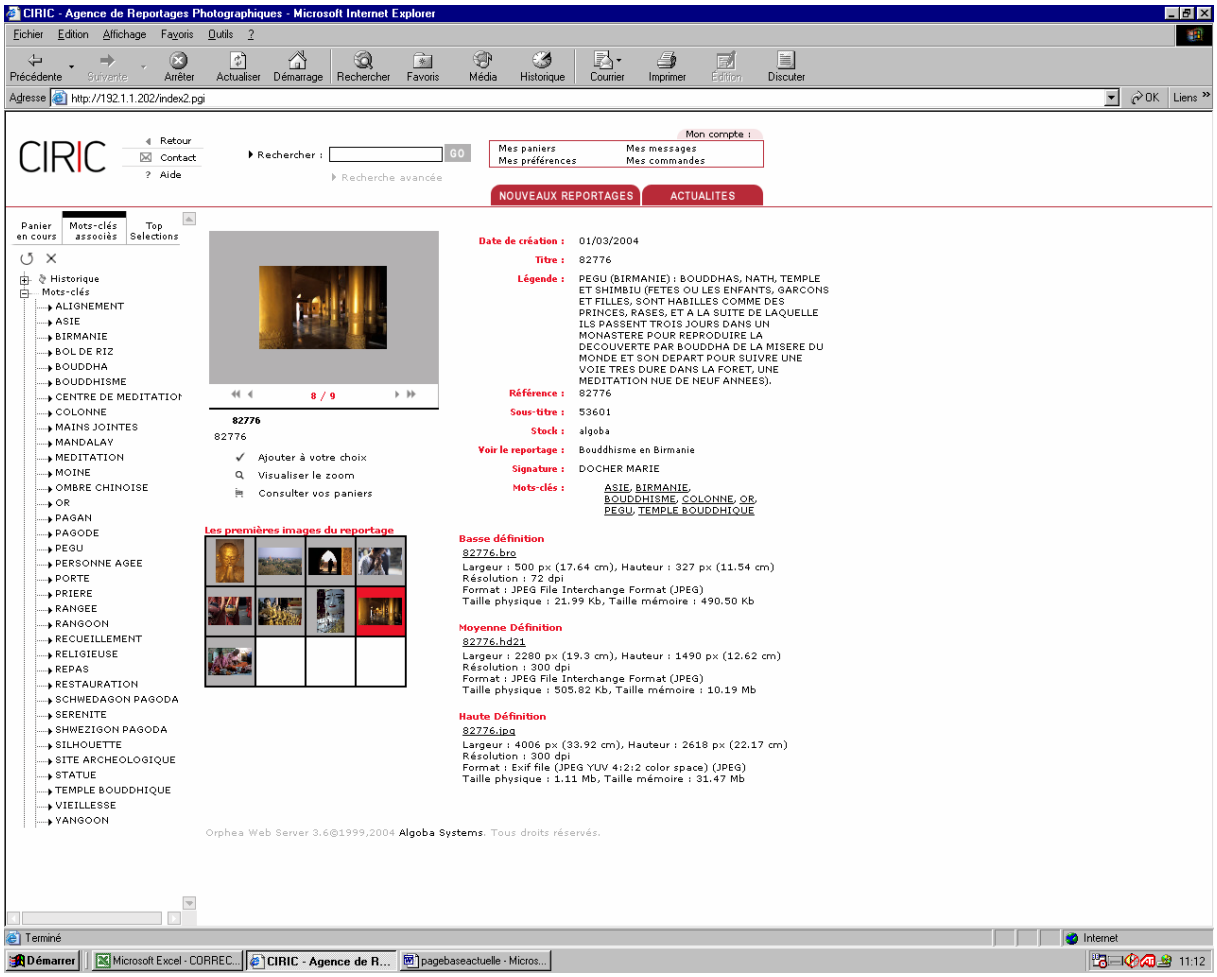
Annexe 6 : Interface de recherche par mots-clés de l'agence Roger Viollet



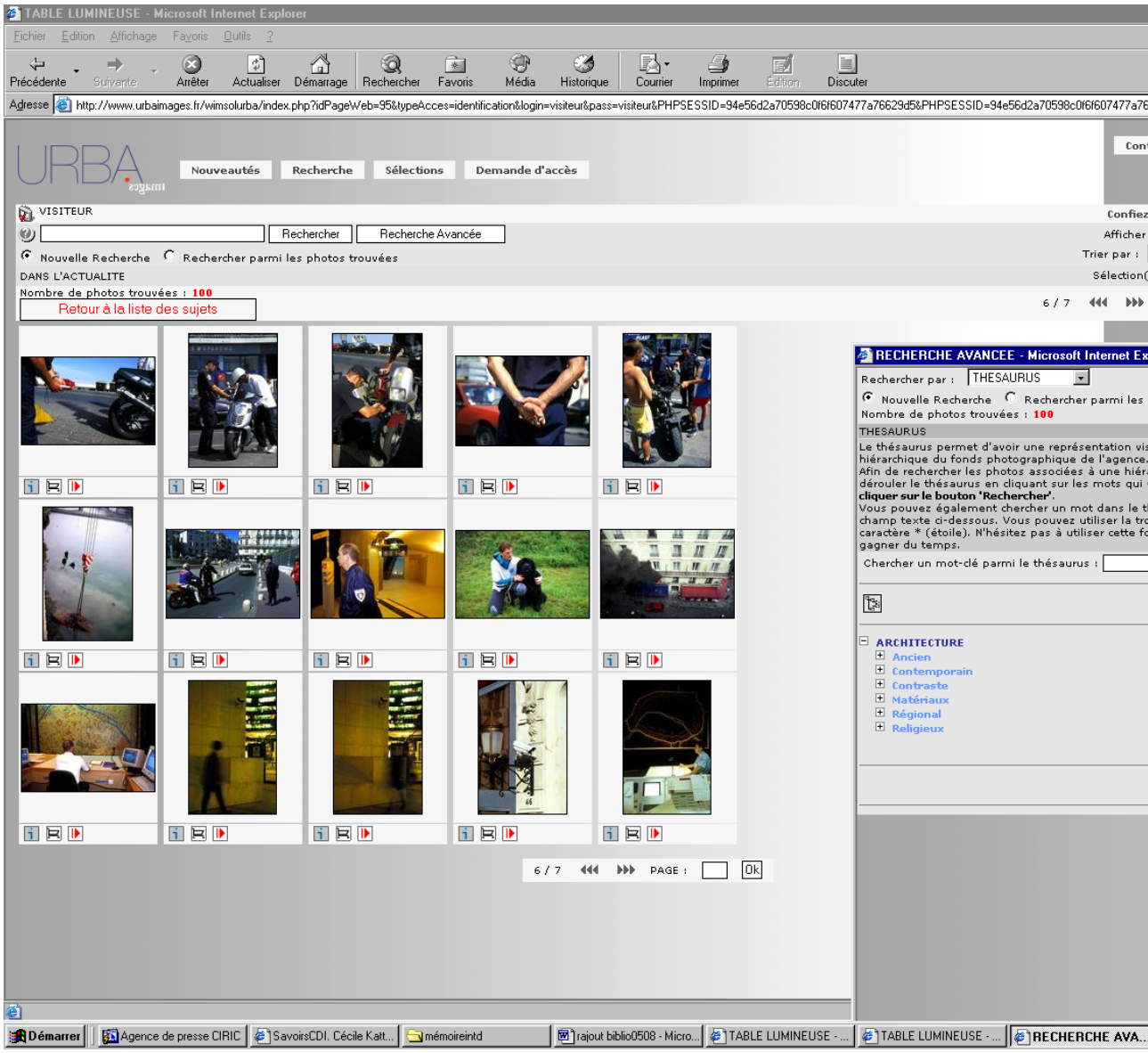
Annexe 7 : Interface de recherche de l'agence Magnum



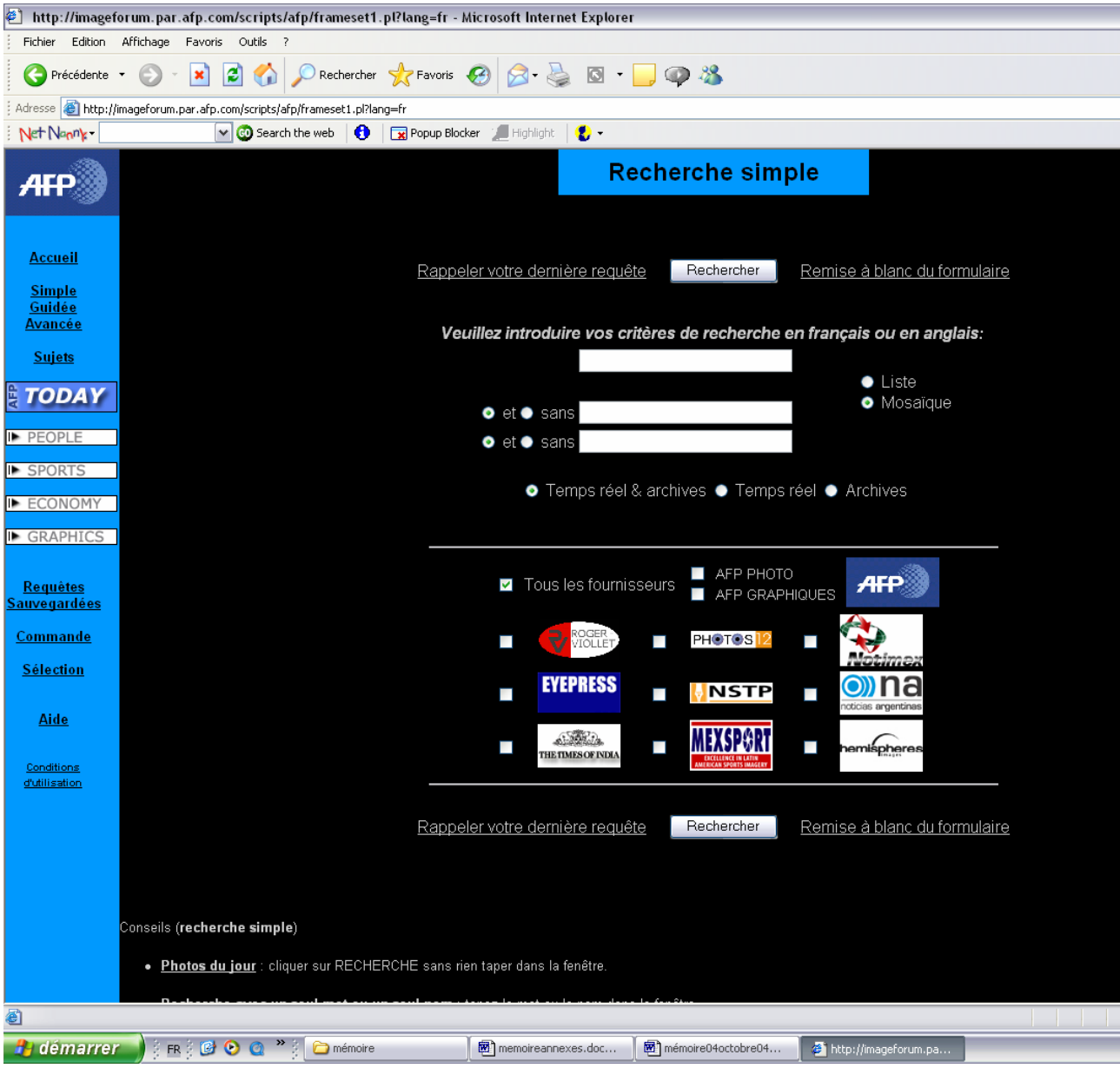
Annexe 8 : Affichage d'une photo avec notice documentaire et mots- clés associés, agence CIRIC



Annexe 9 : Affichage des résultats d'une recherche avec mise en ligne du thésaurus, agence Urba Images



Annexe 10 : Portail d'ImageForum



Annexe 11 : Exemple d'indexation d'images par le contenu (Image Seeker)